GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 20, 2004, 10:09:00 ; Search time 21 Seconds
(without alignments)
2510.141 Million cell updates/sec

Title: US-10-033-245-24
Perfect score: 2868
Sequence: 1 MRLRNGTFLTLLLFCLCAFL.....IQTDSPVWVILSEIFLKKAD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1558	54.3	360	2	T47170	hypothetical prote
2	129	4.5	638	2	E84799	similar to axi 1 p
3	108	3.8	578	2	E64609	DNA polymerase III
4	108	3.8	586	2	B83790	hypothetical prote
5	107.5	3.7	970	2	S01352	type III site-spec
6	107.5	3.7	1288	2	T09908	hypothetical prote
7	106.5	3.7	1273	2	C96767	unknown protein F2
8	106	3.7	1054	2	D70425	conserved hypothet
9	105.5	3.7	1165	2	A70423	valine-tRNA ligase
10	105.5	3.7	6486	2	T31076	tyrocidine synthet
11	105	3.7	2324	1	A29924	acetyl-CoA carboxy
12	104.5	3.6	3770	2	A40889	delta-(L-alpha-ami
13	104	3.6	1039	2	AI2284	translation initia
14	104	3.6	1642	2	T08880	NMDA receptor-bind
15	104	3.6	1818	1	S73852	hypothetical prote
16	103	3.6	2339	2	S41121	acetyl-CoA carboxy
17	102.5	3.6	575	2	G95093	hypothetical prote
18	102.5	3.6	584	2	D97961	hypothetical prote
19	102.5	3.6	3498	2	T22330	hypothetical prote
20	102	3.6	428	2	S09134	gene ND4L intron 1
21	102	3.6	671	2	C83934	methyl-accepting c
22	102	3.6	829	2	F83905	hypothetical prote
23	102	3.6	838	2	A38172	adenylate cyclase
24	102	3.6	1071	2	T04926	starch synthase ho
25	102	3.6	1611	2	A84743	probable myosin he
26	101.5	3.5	916	2	S22864	DNA topoisomerase
27	101	3.5	547	2	S70538	signal-transducing
28	101	3.5	1025	2	T09459	type I site-specif
29	99.5	3.5	1105	2	F71079	hypothetical prote
30	99.5	3.5	4540	2	T30838	cytoplasmic dynein
31	99	3.5	1557	2	T29132	hypothetical prote
32	99	3.5	1870	2	C47521	gag-pol-like fusio
33	99	3.5	4056	2	H96599	protein F14J16.10
34	98.5	3.4	451	1	JE0240	LIM kinase (EC 2.7
35	98.5	3.4	617	2	JC5814	LIM motif-containi
36	98.5	3.4	638	2	JC5813	LIM motif-containi
37	98.5	3.4	783	2	AG3402	polyphosphate kina
38	98.5	3.4	1006	2	S20126	exoribonuclease RA
39	98.5	3.4	1091	2	C95133	exonuclease RexB [
40	98.5	3.4	1091	2	G98001	second chain of ma
41	98	3.4	1025	2	T30322	type I site-specif
42	98	3.4	1025	2	T44802	type I site-specif
43	97.5	3.4	1487	2	S62048	probable membrane
44	97.5	3.4	1839	1	RRWPEM	genome polyprotein
45	97	3.4	697	2	H84791	hypothetical prote

ALIGNMENTS

RESULT 1
T47170
hypothetical protein DKFZp762F216.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47170
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24376
A;Accession: T47170
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-360 <AAA>
A;Cross-references: EMBL:AL162067
A;Experimental source: adult melanoma (MeWo cell line); clone DKFZp762F216
C;Genetics:
A;Note: DKFZp762F216.1

Query Match 54.3%; Score 1558; DB 2; Length 360;
Best Local Similarity 92.6%; Pred. No. 2.4e-112;
Matches 302; Conservative 1; Mismatches 15; Indels 8; Gaps 2;

```
Qy        176 EDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDP 235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 EDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDP 60

Qy        236 KERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMIL 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMIL 120

Qy        296 EFSQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKA 355
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        121 EFSQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDA-HCDRQKA 179

Qy        356 NLRIRFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTL 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 NLRIRFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTL 239

Qy        416 EKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDN 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 EKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLP--- 296

Qy        476 PQSDKEALQEGRTATLRYPRSPDGYL 501
                  | : |   |   | | |
Db        297 ----STTLSQTRRPCRRAAPPPSGTL 318
```

RESULT 2
E84799
similar to axi 1 protein from Nicotiana tabacum [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: E84799
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84799
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-638 <STO>
A;Cross-references: GB:AE002093; NID:g4895186; PIDN:AAD32773.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g37980
A;Map position: 2

Query Match 4.5%; Score 129; DB 2; Length 638;
Best Local Similarity 20.9%; Pred. No. 0.082;
Matches 111; Conservative 76; Mismatches 187; Indels 156; Gaps 27;

```
Qy        103 NGSHRHVLHLPTVFHHLP---HLLAKESSLQPAVRVGQGRTGV----------SVV 145
              :||| |  |    :|| |   : | ::  | | :  | | | |               :
Db         79 SGSHYHHDH---HYHHHPTIRYFLLRKLRL-PFLFDGVGSTAVVGQGWWLCSGRNVGRRI 134

Qy        146 MGIPSVRREVHSYLTDTLHS--LISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALF 203
              :|:  :   |  :|  :| |  ::     ::  : |::  : |  |     |:|||:
Db        135 LGLLMIFVVVSLFLRVSLMSGRVVDHAHRRDLNELVVMRALHEDWSMAQRAMTENV---- 190

Qy        204 PTEIHSGLLEVISPSPHFY--PDFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQS--- 258
                  ::| : | |  :  |:    |:    || | | | | |    :|:::|
Db        191 -------VIEKL-PIPEIWQKPESGNYRQCASRPKNRSRLRRKTN---GYLLVHANGGLN 239

Qy        259 ---KGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILE-----FSQLGFIGKMFKSL 310
                 ||       |:||   ::       |       |          |    |: :   :
Db        240 QMRTGIC------DMVAAAKIMNATLVLPLLDHESFWTDPSTFKDIFDWRHFMNVLKDDV 293

Qy        311 DLSLIVEFILMFYR------DKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFKPS 364
              |  |||::   |         |: |                | :   :   |  : |
Db        294 D---IVEYLPPRYAAMRPLLKAPVSW-------------SKASYYRSEMLPLLKKHKVI 336

Qy        365 LFQHVGT---HSSLAGKIQKLKDK------------DFGKQALRKEHVNPPAEVSTSLK 408
              | |  :   :: |   ||:|: :             ||||  : :   |    ::
Db        337 KFTHTDSRLANNGLPPSIQRLRCRANYQALGYSKEIEDFGKVLVNRLRNNSEPPIAL--- 393

Qy        409 TYQHFTLEKAYLREDFFWAFTP-----AAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKL 463
                |  ||  |      |||       ||:       : ||: |:      |::| ::|
Db        394 ---HLRYEKDML------AFTGCSHNLTAGE------AEELRIMRY-----NVKHWKEKE 433

Qy        464 FNTSVEVLPFDNPQSDKEA---LQEGRTATLRYPRSPDGYLQIGSFYKG---------- 509
               ::    :    | | :||   |:      : || |   |:  |  | |
Db        434 IDSRERRIQGGCPMSPREAAIFLK-----AMGYPSSTTVYIVAGEIYGGNSMDAFREEYP 488

Qy        510 --------VAEGEVDPAFGP----LEALRLSIQTDSPVWVILSEIFLKKA 547
                     | |::| | |    | ||   :  :| |:|   :  : ||
Db        489 NVFAHSYLATEEELEP-FKPYQNRLAALDYIVALESDVFVYTYDGNMAKA 537
```

RESULT 3
E64609
DNA polymerase III gamma and tau subunits - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C;Accession: E64609
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D. Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: E64609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-578 <TOM>
A;Cross-references: GB:AE000584; GB:AE000511; NID:g2313834; PIDN:AAD07767.1; PID:g231384

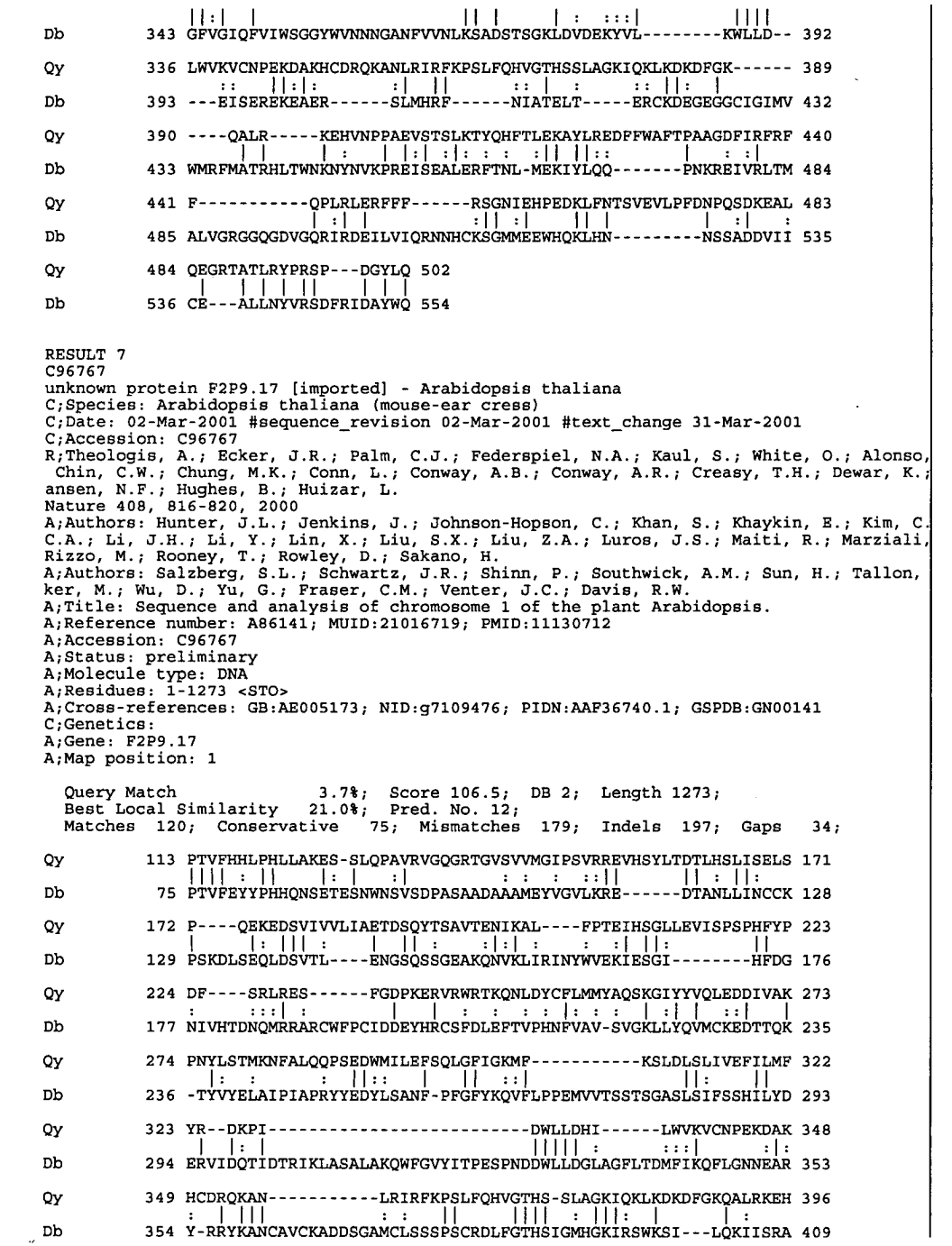
Query Match 3.8%; Score 108; DB 2; Length 578;
Best Local Similarity 19.0%; Pred. No. 2.9;
Matches 115; Conservative 87; Mismatches 177; Indels 226; Gaps 31;

```
Qy         42 EFLALRDR-LHAAE---QESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRL---- 93
              : |||: |  | :|   |||:  :| |:| ||  : | :   :   | |  :  |:
Db          2 QVLALKYRPKHFSELVGQESV--AKTLSLALDNQRLANAYLFSGLRGSGKTSSSRIFARA 59

Qy         94 ---TEDPRLKP----------WNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRT 140
                 | |:  |           |  |  :: :                        |
Db         60 LMCEEGPKAVPCDTCIQCQSALNNHHIDIIEMD--------------------GASNR 97

Qy        141 GVSVVMGI-------PSVRR-------EVHSYLTDTLHSLISELSPQEKEDSVIVVLIAE 186
              |:  |  :     .  ||  |       ||| : |:  ::|:  |   |:  | :  |:|
Db         98 GIDDVRNLIEQTRYKPSFGRYKIFIIDEVHMFTTEAFNALLKTL---EEPPSHVKFLLAT 154

Qy        187 TDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHF----YPD---FSRLRESFGDPKERV 239
              ||:    |                   ::| : ||     |:    | |:      ||:|
Db        155 TDALKLPAT-----------------ILSRTQHFRFKKIPENSVISHLKTIL--EKEQV 194

Qy        240 RWRT--------------------KQNLDYC---------FLMMYAQSKGI---YYVQL 266
              : |                    :| ::||           |: |  : :   ::  |
Db        195 SYETSALEKLAHSGQGSLRDTITLLEQAINYCDNAITESKVAEMLGAIDRSVLEDFFQSL 254

Qy        267 --EDDIVAKPNYLSTMKNFALQQPSEDWMIL--------EFSQLGFIGKMFKSL--DLSL 314
                :|:   |  | : ::|:  :   |: |:         :|  :  | : || :   |||
Db        255 INQDEARLKERY-AILENYETESVLEEMMLFLKAKLLSPDFYSILLIERFFKIIMSSLSL 313

Qy        315 IVE-----FILMFYRDK-----PIDWLLDHILWVKVC----NPEKDAKHCDRQKANLRIR 360
              : |     |:|:  : |         | | || ::     ||        ::  |:  |
Db        314 LKEGANASFVLLLLKMKFKEALKFKALDDAILELEQTPFNQNPSISYNAPKQESKNIEKR 373

Qy        361 FKPSLFQHV-GTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAY 419
               |    : : ||       | :||:          :||:   |      | |       |
Db        374 EKIEQIERIEGTE-----KREKLE---------KKENAETPQTPMLSAKDRIFHNL---- 415

Qy        420 LREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDN---- 475
                              |:  | |  ||          :   :|  ::  : ||:
Db        416 -----------------FKQVQTLVYER--------NYELGAVFEKNIRFIDFDSQTKT 449

Qy        476 -------PQSDKEALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLS 528
                    |||  |:|      |:                  : :   ||  ||  |:::::
Db        450 LTWESLATDKDKELLRE------RFK---------------IVKSIVDGVFGKGESIKIA 488

Qy        529 IQTDS 533
              ::  |
Db        489 LKNHS 493
```

RESULT 4
B83790
hypothetical protein BH1122 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: B83790
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiram Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and s
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: B83790
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-586 <STO>
A;Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BAB04841.1; GSPDB:GN001
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1122

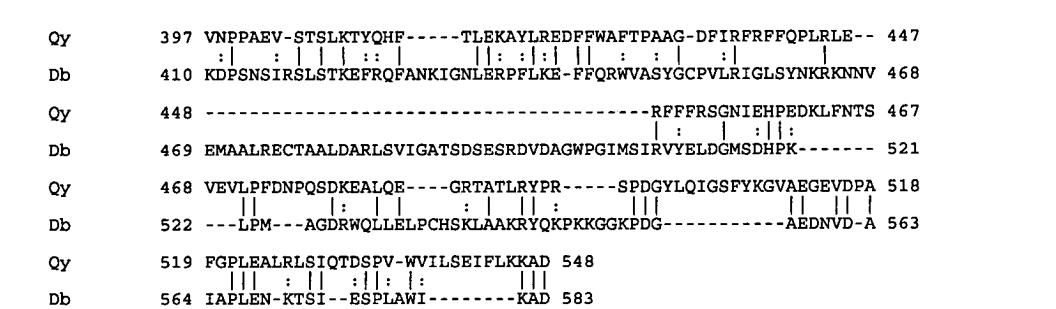
Query Match 3.8%; Score 108; DB 2; Length 586;
Best Local Similarity 20.7%; Pred. No. 3;

```
  Matches  105;  Conservative   72;  Mismatches  147;  Indels  184;  Gaps   27;

Qy          50 LHAAEQES--LKRSKELNL-VLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSH 106
Db         193 VHASTIETWLKQKSQESNLMIIDKQGRFVA----------------TSDPQ-------H 228

Qy         107 RHVLHLPTVFHHLPHLLAKESSLQPAVRVGQG---RTGVSVVMGIPSVR---REVHSYLT 160
Db         229 ESVLYSPYFDQYLPY--------QPTMQTGENDKDEDALEVSMPIPNTEWTLKQVTPY-- 278

Qy         161 DTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPH 220
Db         279 -------SELTEGSRRMAAILVIIGAIMIVVALLVTYSLTKKFTTPIYE-LKNVLSK--- 327

Qy         221 FYPDF-------SRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVA- 272
Db         328 -YPDQDLSGELPSDYRNEFGELFEGYKGLIRRNNE-----LYRSLHRQYKRQREAEIKAL 381

Qy         273 ----KPNYL-STMKNFALQQPSEDWMILE------FSQLGFIGKMFKSLDLSLIVEFILM 321
Db         382 QANINPHFLYNTLDQL-------NWMAIEKGDRTMSHMLENLGQMLR-IGLS-NGESILP 432

Qy         322 FYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQK 381
Db         433 L--EKEVEYLKYYL-----------------------------------------KIQK 448

Qy         382 LKDKDFGKQALRKEHVNPPAEVSTSLKTY--QHFTLEKAYLREDFFWAFTPAAGDFIRFR 439
Db         449 LK---------MGERLTYTVDVPVMLYRYLIPKLTLQ-PFVENCFIHGFQDGRDGEVVIR 498

Qy         440 F-----------------FQPL-------------------RLERFFFRSGNIEHPEDKL 463
Db         499 AMEDQNHIVMTIQDNGIGFQPVTTKRSKLDMGGYGIRNVMERLDVYFDRYASVHVTGDEQ 558

Qy         464 FNTSVEVLPFDNPQ-SDKEALQEGRTAT 490
Db         559 GGTLVKII---IPKVLNKERLLDGRSIT 583


RESULT 5
S01352
type III site-specific deoxyribonuclease (EC 3.1.21.5) EcoP1 chain res - phage P1
N;Alternate names: type III restriction enzyme EcoP1 chain res
C;Species: phage P1
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 08-Oct-1999
C;Accession: S01352
R;Huembelin, M.; Suri, B.; Rao, D.N.; Hornby, D.P.; Eberle, H.; Pripfl, T.; Kenel, S.; E
J. Mol. Biol. 200, 23-29, 1988
A;Title: Type III DNA restriction and modification systems EcoP1 and EcoP15. Nucleotide
A;Reference number: S01351; MUID:88245189; PMID:2837577
A;Accession: S01352
A;Molecule type: DNA
A;Residues: 1-970 <HUE>
A;Cross-references: EMBL:X06287; NID:g15138; PIDN:CAA29615.1; PID:g15140
C;Keywords: hydrolase

  Query Match             3.7%;  Score 107.5;  DB 2;  Length 970;
  Best Local Similarity   19.1%;  Pred. No. 6.9;
  Matches  105;  Conservative   76;  Mismatches  161;  Indels  207;  Gaps   28;

Qy          54 EQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLP 113
Db         477 EINEILHDKELLLSLDNPRRFIPSKWTLREG----------------WDN--------P 511

Qy         114 TVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYL-------------- 159
Db         512 NVF-QICKLRSSGSTTSKLQEVGRG------------LRLPVNEYMCRVKDRNFTLKYYV 558

Qy         160 ----TDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLL--- 212
Db         559 DFTEKDFVDSLVKEVNESSFKERV--------PSKFTQELKEQIRAQYP-ELSSRALMNE 609

Qy         213 ----EVISPSPHFYPD--FSRLRESF----------------GDPKERVRWRTKQNLDYC 250
                     |:|  : :|     :|||:  :                   | | | : |  :  :
Db         610 LFNDEIIDDNDNFKDSDAYSRLKSKYPAAFPIGVKPGKIKKATDGKRRTKMRVGKFSELK 669

Qy         251 FLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLGF---IGKMF 307
Db         670 ELWELINQKAV----IEYKINSENEFLSIFKSFMLEETE------RFTKSGVHTRIDKIY 719

Qy         308 KSLDLSLIVEFI----------LMFYRDKPIDWLLDHI---LWVKVCNPEKDAKH---CD 351
Db         720 IHNDMAMSKSIVSDDDDFAKLNTMSYRE-----FLDNLSQTIFVK-----HDTLHKVFCD 769

Qy         352 -RQKANL-------RIRFKPSLFQHVGTHSS----------LAGKIQKLK---------- 383
Db         770 IKDTINITEYLNIQTIRKIKSGFSKYLLNNSFNKFSLGYNLISGSIHPTKFTNADGKPLD 829

Qy         384 ---DKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYLRE----------------- 422
Db         830 EVLSSDLG---VLQDNSKAPLDTYLFEEVFYDSELERRNITDREIQSVVVFSKIPKNSIK 886

Qy         423 -----------DFFWAFTPAAGDFIRFRF-------FQPLRLERFFFRSGNIEHPEDKLF 464
Db         887 IPVAGGYTYSPDFAYVVKTAEGDYLNFIIETKNVDSKDSLRLE----EKKKIEHAQ-ALF 941

Qy         465 N---TSVEV 470
Db         942 NQISQSVKV 950


RESULT 6
T09908
hypothetical protein T22A6.280 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T09908
R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Maye
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16896
A;Accession: T09908
A;Molecule type: DNA
A;Residues: 1-1288 <BEV>
A;Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.280
A;Experimental source: cultivar Columbia; BAC clone T22A6
C;Genetics:
A;Gene: ATSP:T22A6.280
A;Map position: 4
A;Introns: 18/2; 31/3; 68/2; 89/3; 120/2; 170/3; 272/3; 313/2; 378/3; 408/2; 453/2; 507/3
1151/3; 1182/3; 1221/2

  Query Match             3.7%;  Score 107.5;  DB 2;  Length 1288;
  Best Local Similarity   20.1%;  Pred. No. 10;
  Matches  113;  Conservative   88;  Mismatches  168;  Indels  193;  Gaps   31;

Qy          51 HAAEQESLK----RSKELNLVLDEIK--RAVSERQALRDGDGNRTWGRLTEDPRLK-PWN 103
Db          76 HSSKQGALQTTFVKSGDAYVVILELRDPRVRAIEFVLKDGSHNRWLRQHNGNFRVEIPWN 135

Qy         104 GSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGV-SVVMGIPSVRREVHSYL--- 159
Db         136 DLHAH--------HRIPKTLIERRAHKIWDRKGRPQSSAREQQIDYDNAVRELHAELARG 187

Qy         160 -------------------TDTLHSLISE-------------LSPQEKEDSVIVVLIAE 186
Db         188 ISLDELQANSTVPVEKEETSEPHHTMIQSYRRKHDVQKWLQKYTEPINRSGSVKSSALAE 247

Qy         187 TDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRE--SFGDPKERVRWRTK 244
Db         248 LSKR--SVGQENLVSQKSFHVRNYEITIISRSSK-----GMLREIVAYGLPR---TWQVQ 297

Qy         245 QNLDYCFLMMYAQSKGIYYVQLEDD----IVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
Db         298 Q-----FSIGESQS------HLQESGCTFILIYLNLLTHQIFFTL----EKFIDINLKRG 342

Qy         301 GFIGKMF-------------------KSLD-----LSLIVEFILMFYRDKPIDWLLDHI 335
```



RESULT 7
C96767
unknown protein F2P9.17 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96767
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C96767
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1273 <STO>
A;Cross-references: GB:AE005173; NID:g7109476; PIDN:AAF36740.1; GSPDB:GN00141
C;Genetics:
A;Gene: F2P9.17
A;Map position: 1

Query Match 3.7%; Score 106.5; DB 2; Length 1273;
Best Local Similarity 21.0%; Pred. No. 12;
Matches 120; Conservative 75; Mismatches 179; Indels 197; Gaps 34;



RESULT 8
D70425
conserved hypothetical protein aq_1442 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 15-Sep-2003
C;Accession: D70425
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: D70425
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1054 <AQF>
A;Cross-references: GB:AE000740; NID:g2983826; PIDN:AAC07394.1; PID:g2983839; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1442
C;Superfamily: oxygen sensor diguanylate cyclase/c-di-GMP phosphodiesterase

Query Match 3.7%; Score 106; DB 2; Length 1054;
Best Local Similarity 19.6%; Pred. No. 10;
Matches 120; Conservative 86; Mismatches 183; Indels 224; Gaps 32;

```
Db        694 ------LLKEIAVRLEKTFRRGDII-----AKGDGDEFLIFAYPLKKKDDVLSLIEKIRQ 742

Qy        439 RFFQPLRLE----RFFFRSGNIEHPED-----KLFNTSV------------EVLPFDNPQ 477
              :| :|| ::    |  |  | : :| |     :|:  ::           ||: ||  :
Db        743 KFSKPLTIDGKKIRLSFYIGVVTYPYDGRTFSELYYRAITAVKEAKRRGAGEVVLFDE-E 801

Qy        478 SDKEA--LQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPV 535
               |||   | ||   ||             | | |     | |:  ||     |  :
Db        802 LDKEVERLIEGE-KLLR--------------KAVEENLFTFHFQPIYRLR-----DMKI 840

Qy        536 WVILSEIFLKKAD 548
              : : : : :|: |
Db        841 FSLEALVRIKEGD 853


RESULT 9
A70423
valine-tRNA ligase (EC 6.1.1.9) - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 03-Jun-2002
C;Accession: A70423
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: A70423
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1165 <AQF>
A;Cross-references: GB:AE000739; NID:g2983813; PIDN:AAC07375.1; PID:g2983819; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: valS
C;Superfamily: valine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

  Query Match             3.7%;  Score 105.5;  DB 2;  Length 1165;
  Best Local Similarity   19.1%;  Pred. No. 13;
  Matches   66;  Conservative   58;  Mismatches  103;  Indels  119;  Gaps   17;

Qy        174 EKEDSVIVVLIAETDSQYTSAV-----------TENIKALFPTE--------IHSGLLEV 214
              |:|  |:        |: ::||:           ||::| |:||:        |   :  :
Db        719 EQERDVL-------DTWFSSALWPFGVFGWPESTEDLKNLYPTDLLVTGFDIIFFWVARM 771

Qy        215 ISPSPHFYPDF--------SRLRESFG-----------DPKERVRWRTKQNLDYCFLMMY 255
              |    ||  |          : :|: :|           || : :        | :   ::
Db        772 IMMGTHFMKDIPFYDVYVHALVRDKYGRKMSKTIGNVIDPLDIIERYGADALRFTLAILT 831

Qy        256 AQSKGI--------------------YYVQLE--DDIVAKPNYLSTMKNFALQQPSEDW 292
              | : |                    || :   :| :|:  |:: :|       | : |
Db        832 VQGRDIKLAEEKFEGYKHFANKIWNAARYVLMNTPEDFIARIPYMAPLK------PEDKW 885

Qy        293 MILEFSQLG-FIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHI---LWVKV--CNPEKD 346
              :| : ::    : |  ::   |     |  |:     || ::     :: |    |:|:
Db        886 IITKLNETAEEVNKALENYQYSQAAHAIYEFFWSDYCDWYIEFTKERIYKKAPEDNEEEK 945

Qy        347 AK------------HCDRQKANLRI--RFKPSLFQHV--------GTHSSLA-------- 376
              ||            |   :|| |||   | | : : :        |   |||
Db        946 AKVENERTTALYTLHYVLEKA-LRILHPFMPYITEELWHKLPNAEGESISLAEFPQKNED 1004

Qy        377 -------GKIQKLKDKDFGKQALRKE-HVNPPAEVSTSLKTYQHFT 414
                    |:::||:     :|:| :  : |  ::  | ||   |:
Db       1005 EIYEEDKQKVERLKEIISAIRAIRSDLQIKPSEKIKASFKTESEFS 1050


RESULT 10
T31076
tyrocidine synthetase 3 - Brevibacillus brevis
C;Species: Brevibacillus brevis
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C;Accession: T31076
R;Mootz, H.D.; Marahiel, M.A.
J. Bacteriol. 179, 6843-6850, 1997
A;Title: The tyrocidine biosynthesis operon of Bacillus brevis: Complete nucleotide seque
A;Reference number: Z20969; MUID:98012987; PMID:9352938
A;Accession: T31076
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6486 <MOO>
A;Cross-references: EMBL:AF004835; NID:g2623770; PID:g2623773; PIDN:AAC45930.1
C;Genetics:
A;Gene: tycC
C;Function:
A;Pathway: tyrocidine biosynthesis
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;510-950/Domain: acetate-CoA ligase homology <ACL1>
F;968-1036/Domain: acyl carrier protein homology <ACP1>
F;1546-1987/Domain: acetate-CoA ligase homology <ACL2>
F;2005-2073/Domain: acyl carrier protein homology <ACP2>
F;2583-3025/Domain: acetate-CoA ligase homology <ACL3>
F;3043-3111/Domain: acyl carrier protein homology <ACP3>
F;3621-4060/Domain: acetate-CoA ligase homology <ACL4>
F;4078-4146/Domain: acyl carrier protein homology <ACP4>
F;4656-5104/Domain: acetate-CoA ligase homology <ACL5>
F;5122-5190/Domain: acyl carrier protein homology <ACP5>
F;5702-6147/Domain: acetate-CoA ligase homology <ACL6>
F;6165-6233/Domain: acyl carrier protein homology <ACP6>
F;1000,2037,3075,4110,5154,6197/Binding site: phosphopantetheine (Ser) (covalent) #status

  Query Match             3.7%;  Score 105.5;  DB 2;  Length 6486;
  Best Local Similarity   21.5%;  Pred. No. 1.6e+02;
  Matches  113;  Conservative   73;  Mismatches  216;  Indels  123;  Gaps   27;

Qy         50 LHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWG------RLTED---PRLK 100
              |:| :| | | :  :  : : | | |:|   ::    ||:        |  |   | ::
Db       4309 LYAGKQLSDLRIQYKDFAVWQTKLAQSDRFQKQEDFWTRTFAGEIPLLNLPHDYPRPSVQ 4368

Qy        101 PWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLT 160
               ::|     : | |  |||   | | ::           || :: |           |
Db       4369 SFDGD---TVALGT-GHHLLEQLRKLAA----------ETGTTLFM-----------VLL 4403

Qy        161 DTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPH 220
               | |:|: : ||:      :::    :  : |  | |  :|   :   |    : |
Db       4404 AAYHVLLSKYAGQEE------IVVGTPIAGRSHADVERIVGMFVNTL--ALKNTAAGSLS 4455

Qy        221 FYPDFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTM 280
              |       |  | |:          :: || |  :      :  :|:  |:   | :  ||
Db       4456 F-------RAFLEDVKQNA-LHAFEHQDYPFEHL------VEKLQVRRDLSRNPLF-DTM 4500

Qy        281 KNFALQQPSE----DWMILEFSQLGFIGKMFKSLDL-----SLIVEF--ILMFYRDKPID 329
              :  | : :|     |  :  :   | | |   |||        |:|:|      :  : :|
Db       4501 FSLGLAESAEGEVADLKVSPYPVNGHIAKFDLSLDAMEKQDGLLVQFSYCTKLFAKETVD 4560

Qy        330 WLLDH---ILWVKVCNPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQKLKDKD 386
              |  |   :|      :|       | : | : :  |    :|:  || || |     ||
Db       4561 RLAAHYVQLLQTITADP-------DIELARISVLSKAET-EHM-LHSFLATKTAYPTDKT 4611

Qy        387 FGKQALRKEHV-NPPAEVS----TSLKTYQHFT-----LEKAYLREDF------------ 424
              |  | | :| |   | |::       |||         | :   |:
Db       4612 F--QKLFEEQVEKTPNEIAVLFGNEQLTYQELNAKANQLARVLRRKGVKPESTVGILVDR 4669

Qy        425 -------FWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNT-----SVEVLP 472
                    |   | | |:       ||  : |      :|  | ||  |     |  |
Db       4670 SLYMVIGMLAVLKAGGTFVPIDPDYPLERQAFM-----LEDSEAKLLLTLQKMNSQVAFP 4724

Qy        473 FDNPQSDKEALQEGRTATLRYPRSPD--GYLQIGSFYKGVAEGEV 515
              ::    | | : :  |  | :   |:   |:   |    |  :| |
Db       4725 YETFYLDTETVDQEETGNLEHVAQPENVAYIIYTSGTTGKPKGVV 4769


RESULT 11
A29924
acetyl-CoA carboxylase (EC 6.4.1.2), hepatic - chicken
C;Species: Gallus gallus (chicken)
```

```
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Feb-2002
C;Accession: A29924; A29337; A27903
R;Takai, T.; Yokoyama, C.; Wada, K.; Tanabe, T.
J. Biol. Chem. 263, 2651-2657, 1988
A;Title: Primary structure of chicken liver acetyl-CoA carboxylase deduced from cDNA seq
A;Reference number: A29924; MUID:88139305; PMID:2893793
A;Accession: A29924
A;Molecule type: mRNA
A;Residues: 1-2324 <TAK1>
A;Cross-references: GB:J03541; NID:g211567; PIDN:AAA48701.1; PID:g211568
R;Takai, T.; Wada, K.; Tanabe, T.
FEBS Lett. 212, 98-102, 1987
A;Title: Primary structure of the biotin-binding site of chicken liver acetyl-CoA carbox
A;Reference number: A91375; MUID:87106011; PMID:2879745
A;Accession: A29337
A;Molecule type: mRNA
A;Residues: 493-820 <TAK2>
A;Cross-references: GB:X05019; NID:g63021; PIDN:CAA28675.1; PID:g1334694
A;Accession: A27903
A;Molecule type: mRNA
A;Residues: 493-552,554-783,'RSPS',789-820 <TAK3>
C;Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
C;Keywords: biotin binding; ligase; liver
F;120-620/Domain: biotin carboxylase homology <BCH>
F;747-819/Domain: lipoyl/biotin-binding homology <LPB>
F;786/Binding site: biotin (Lys) (covalent) #status predicted

  Query Match             3.7%;  Score 105;  DB 1;  Length 2324;
  Best Local Similarity   17.5%;  Pred. No. 39;
  Matches  106;  Conservative   82;  Mismatches  202;  Indels  216;  Gaps   25;

Qy         64 LNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLK---------------------- 100
              : |:||  ||     ||:  |     ||  :|:|:|
Db        190 VELILDIAKRI--PVQAVWAG-----WGHASENPKLPELLHKNGIAFMGPPSQAMWALGD 242

Qy        101 ----------------PWNGS-----------HRHVLHLPTVFHHLPHLLAKESSLQPAV 133
                             |||||           : :|::|   :   ::   :  |: |
Db        243 KIASSIVAQTAGIPTLPWNGSGLRVDWQENDLQKRILNVPQELYEKGYVKDADDGLRAAE 302

Qy        134 RVG--------QGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVIVVLIA 185
               ||        :|  |  :         |:|::   |   :|  ::  :     : |: :|
Db        303 EVGYPVMIKASEGGGGKGI--------RKVNN--ADDFPNLFRQVQAEVPGSPIFVMRLA 352

Qy        186 ETD---------SQYTSAVTENIKALFPTEI-----HSGLLEVISPSPHFYPDFSRLRES 231
              :           || :|:     :||  :       |  ::|     |       | : :
Db        353 KQSRHLEVQILADQYGNAI-----SLFGRDCSVQRRHQKIIEEAPASIATSVVFEHMEQC 407

Qy        232 FGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLE---------DDIVAKPNYLSTMKN 282
                  : | : :   ::|    :|:|    |:::|           ::|| |  :
Db        408 AVKLAKMVGYVSAGTVEY----LYSQDGSFYFLELNPRLQVEHPCTEMVADVNLPAAQLQ 463

Qy        283 FALQQPSEDWMILEFSQLGFIGKMFKSLDLSLIVEFILMF----YRDKPIDW-------- 330
               |: |                     |  | :  :|:    : |  ||:
Db        464 IAMGIP----------------------LHRIKDIRVMYGVSPWGDGSIDFENSAHVPC 500

Qy        331 LLDHILWVKVC--NPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQKLKDKDFG 388
                 |::  ::   ||::  |       |  |   ::: :     : || : :  |  ||
Db        501 PRGHVIAARITSENPDEGFKPSSGTVQELNFRSNKNVWGYFSV--AAAGGLHEFADSQFG 558

Qy        389 K----QALRKEHVNPPAEVSTSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPL 444
                   |:| ::          |       |     || :              :    ||
Db        559 HCFSWGENREEAISNMVVALKELSIRGDFRTTVEYLIK------------LLETESFQQN 606

Qy        445 R-----LERFFFRSGNIEHPE-------------DKLFNTSV----------EVLPFDNP 476
              |     |:|       | |:             |  |  ||          :|||
Db        607 RIDTGWLDRLIAEKVQAERPDTMLGVVCGALHVADVSFRNSVSNFLHSLERGQVLPAHTL 666

Qy        477 QS--DKEALQEGRTATLRYPR-SPDGYLQI----------------------GSFYKGV 510
               :  | | : |||    |:  | ||: |: |                      || |
Db        667 LNTVDVELIYEGRKYVLKVTRQSPNSYVVIMNSSCVEVDVHRLSDGGLLLSYDGSSYTTY 726

Qy        511 AEGEVD 516
               : |||
Db        727 MKEEVD 732


RESULT 12
A40889
delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase - Emericella nidulans
N;Alternate names: ACV synthetase
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 03-Nov-2000
C;Accession: A40889; S16466
R;MacCabe, A.P.; van Liempt, H.; Palissa, H.; Unkles, S.E.; Riach, M.B.R.; Pfeifer, E.; \
J. Biol. Chem. 266, 12646-12654, 1991
A;Title: delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from Aspergillus nid
thway.
A;Reference number: A40889; MUID:91286299; PMID:2061333
A;Accession: A40889
A;Molecule type: DNA
A;Residues: 1-3770 <MAC>
A;Cross-references: GB:X54853; NID:g2318; PIDN:CAA38631.1; PID:g2319
A;Note: the sequence of residues 3129-3148 and the corresponding nucleotide sequence are
C;Genetics:
A;Gene: acvA
C;Superfamily: alpha-aminoadipyl-cysteinyl-valine synthetase; acetate-CoA ligase homology
C;Keywords: antibiotic biosynthesis; carrier protein; penicillin biosynthesis; phosphopar
F;365-831/Domain: acetate-CoA ligase homology <ACL1>
F;848-918/Domain: acyl carrier protein homology <ACP1>
F;1458-1915/Domain: acetate-CoA ligase homology <ACL2>
F;1931-2001/Domain: acyl carrier protein homology <ACP2>
F;2539-3001/Domain: acetate-CoA ligase homology <ACL3>
F;3018-3086/Domain: acyl carrier protein homology <ACP3>
F;882,1965,3050/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             3.6%;  Score 104.5;  DB 2;  Length 3770;
  Best Local Similarity   19.7%;  Pred. No. 86;
  Matches  108;  Conservative   79;  Mismatches  195;  Indels  167;  Gaps   24;

Qy         39 YQREF---LALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTE 95
              | ||    : : : |       |: :| |:  :  ::   | : |
Db       1352 YARELFDEIVISELLQMVRDTLLQVAKHLDDPVRSLEYLSSAQMA--------------- 1396

Qy         96 DPRLKPWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVR-RE 154
                :|  || :       |    |   :  ||:: :|         |:||    |:  |:
Db       1397 --QLDAWNATD---AEFPDTTLHA--MFEKEAAQKP--------DKVAVVYEQRSLTYRQ 1441

Qy        155 VHSYLTDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEV 214
              ::       | | |::||  | :|:| ::: :::    :     | |:: |    |
Db       1442 LNERANRMAHQLKSDISP--KPNSIIALVVDKSEHMIAT-----ILAVWKT---GGAYVP 1491

Qy        215 ISPSPHFYPD--------------------FSRLRESFGDP----KERVRWRTKQNLDY 249
              | |    |||                    ||::|  |:     :  :  :|  |
Db       1492 IDPE---YPDDRIRYILEDTSAIAVISDACYLSRIQELAGESVRLYRSDISTQTDGNWSV 1548

Qy        250 CFLMMYAQSKGIYYVQLEDDIVAKPN------------YLSTMKNFALQQPSEDWMILEF 297
                  : |  : |:         ||              :|  | | |:  ::| :|| |
Db       1549 SNPAPSSTSTDLAYIIYTSGTTGKPKGVMVEHHGVVNLQISLSKTFGLRD-TDDEVILSF 1607

Qy        298 SQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANL 357
              |   |                      |  :: : | ||  :       ||   |:::
Db       1608 SNYVF---------------------DHFVEQMTDAILNGQTLVMLNDAMRSDKER--- 1642

Qy        358 RIRFKPSLFQHVGTH--SSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTL 415
                 |:|:: |:  :  |:|                        : |: :   |:
Db       1643 -------LYQYIETNRVTYLSG-------------------------TPSVISMYEFSR 1669

Qy        416 EKAYLR--EDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPF 473
               | :||  :    ||:    | ||   || | :  :          | :    |   : ||
Db       1670 FKDHLRRVDCVGEAFSQPVFDQIR-DTFQGLIINGY--------GPTEISITTHKRLYPF 1720

Qy        474 DNPQSDKEALQEGRTAT--------LRYPRSPDGYLQIGSFYKGVAEGEVD-PAFGPLEA 524
              ::||   |:   :|          | |    | | :|   :||| |  : |
Db       1721 PERRTDKSIGQQIGNSTSYVLNADMKRVPIGAVGELYLGG--EGVARGYHNRPEVTAERF 1778

Qy        525 LRLSIQTDS 533
```

```
                ||   ||||
Db       1779 LRNPFQTDS 1787


RESULT 13
AI2284
translation initiation factor IF-2 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AI2284
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2284
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1039 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75531.1; PID:g17132966; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: infB
C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu homo

  Query Match             3.6%;  Score 104;  DB 2;  Length 1039;
  Best Local Similarity   20.6%;  Pred. No. 14;
  Matches  100;  Conservative   78;  Mismatches  158;  Indels  150;  Gaps   26;

Qy          62 KELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLPH 121
               ||| :   |:: |   | :|           ::||   :    |   |:|  | |   : |
Db         501 KELEI---EVETAEPEAEA----------RKVTEMIEV----GDLEHLLRRPPVVTIMGH 543

Qy         122 LLAKESSLQPAVR---VGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDS 178
               :   :::|  ::|   |  |  |          : : : :|  | :|         : ::
Db         544 VDHGKTTLLDSIRKTKVAAGEAG--------GITQHIGAYHVDIVH---------DGKEQ 586

Qy         179 VIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKER 238
               || |       :|:      :    |:|   :| | :                    :
Db         587 QIVFLDTPGHEAFTAMRARGARV---TDI--AVLVVAA------------------DDG 622

Qy         239 VRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFS 298
               || :| : :       :||: |:        ||   | :   |  |  ||  | :  | :
Db         623 VRPQTVEAIS------HAQAAGV-------PIVVAINKID--KEGA--QP--DRVKQELT 663

Qy         299 QLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVK-----VCNPEKDAK----- 348
               | |   : :    : : |  |    | : :| ||: || |        ||:::|:
Db         664 QYGLTSEEWGGETIMVPVSAI----RGENLDTLLEMILLVAEVGELSANPDRNARGTVIE 719

Qy         349 -HCDRQKANLRIRFKPSLFQHVG---THSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVS 404
                | |: |  :      :   |||      |  ||:: : | | |    |:  :  |:
Db         720 AHLDKAKGAVATLLIQNGTLHVGDILLAGSAFGKVRAMVD-DRG----RRVDIAGPS--- 771

Qy         405 TSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLF 464
                     | :|   | :       |||||                          | ::|
Db         772 --------FAVEVLGLSD------VPAAGD------------------------EFEVF 792

Qy         465 NTSVEVLPFDNPQSDKEALQ---EGR-TATLRYPRSPDGYL-QIGSFYKGVAEGEVDPAF 519
               :   |     : ::||: |    :|| | |    :: :| | ::    ||  :| |:
Db         793 DNEKEARALASDRADKQRLSRLLQGRVTLTTLSAQAQEGELKELNLILKGDVQGSVEAIV 852

Qy         520 GPLEAL 525
               | |: :
Db         853 GSLKQI 858


RESULT 14
T08880
NMDA receptor-binding protein yotiao - human
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: T08880
R;Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A;Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
A;Reference number: Z16511; MUID:98151389; PMID:9482789
A;Accession: T08880
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1642 <LIN>
A;Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068
C;Genetics:
A;Map position: 7q21-22
C;Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

  Query Match             3.6%;  Score 104;  DB 2;  Length 1642;
  Best Local Similarity   20.2%;  Pred. No. 28;
  Matches   87;  Conservative   56;  Mismatches  104;  Indels  184;  Gaps   21;

Qy          27 ALSGQKGDVVDVYQREFLALRDRLHAAEQESLK------RSKELNLVLDEIKRAVSER-- 78
               ||   | :::   : :   |: ::|  | :::|         | |:|:: ::::||||
Db        1154 ALCSLKEELIFAQEEKIKELQ-KIHQLELQTMKTQETGDEGKPLHLLIGKLQKAVSEECS 1212

Qy          79 ---QAL-----------------RDGDGNRTWGRLTEDPRLKPWN------GSHRHVLH 111
                  | |                 | : :  :    ||| |: :          : | |
Db        1213 YFLQTLCSVLGEYYTPALKCEVNAEDKENSGDYISENEDPELQDYRYEVQDFQENMHTL- 1271

Qy         112 LPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRRE------VHSYLTD---- 161
               |  |       ||  :: |       ||    |: :  |  :: :|      :|| :|:
Db        1272 LNKVTEEYNKLLVLQTRLSKI--WGQQTDGMKLEFGEENLPKEETEFLSIHSQMTNLEDI 1329

Qy         162 ------------------------TLHSLISELSPQEKE---------------------- 176
                                      | |||| |  | ||
Db        1330 DVNHKSKLSSLQDLEKTKLEEQVQELESLISSLXQQLKETEQNYEAEIHCLQKRLQAVSE 1389

Qy         177 ---------DSVIVVLIAETDSQYT----SAVTENIKALFPTEIHSGLLEVISPSPHFYP 223
                        |||:   | |:|:| |     | | :||          | :|
Db        1390 STVPPSLPVDSVV---ITESDAQRTMYPGSCVKKNI---------DGTIE---------- 1427

Qy         224 DFSRLRESFGDPKER--VRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIV---------- 271
                 ||      ||  :|   |:   ||                   | |||:::
Db        1428 -FS---GEFGVKEETNIVKLLEKQ---------------YQEQLEEEVAKVIVSMSIAF 1467

Qy         272 -----------AKPNYLSTMKNFALQQPSEDW---MILEFSQLGFIGKMFKSLDLSLIVE 317
                        | |  |: :  |: |  : :   | |    |:||  : |:::|:    |
Db        1468 AQQTELSRISGGKENTASSKQAHAVCQQEQHYFNEMKLSQDQIGF--QTFETVDVKFKEE 1525

Qy         318 FILMFYRDKPI 328
               |       ||:
Db        1526 F-------KPL 1529


RESULT 15
S73852
hypothetical protein MG218 homolog F10_orf1818 - Mycoplasma pneumoniae (strain ATCC 29342
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: S73852
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73852
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1818 <HIM>
A;Cross-references: EMBL:AE000051; GB:U00089; NID:g1674211; PIDN:AAB96174.1; PID:g167422
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: Mycoplasma genitalium hypothetical protein MG218

  Query Match             3.6%;  Score 104;  DB 1;  Length 1818;
  Best Local Similarity   19.7%;  Pred. No. 32;
```

Matches 71; Conservative 55; Mismatches 122; Indels 112; Gaps 15;

```
Qy          49 RLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRH 108
               :||  :|  : : :||  : : : | :|     |:
Db        1490 KLHLQKQSIISKGQELKEIKERVSRDISHTNKQRE------------------------ 1524

Qy         109 VLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVH---SYLTDTLHS 165
                     |  || :   ||                 :    ||::   | ||  : :
Db        1525 ---------ELNSLLHQNKLLQ--------------KNLAERERBINNKDSLLTQKIQT 1560

Qy         166 LISELSPQEKEDSVIVVL--IAETDSQYTSAVT----------ENIKALFPT--EIHSG 210
                  :||  |||  :: :|  :   : || : :|           : | |||   :|:
Db        1561 AKQKLS--EKEARILKLLEKMRAVEQQYQAEITRLKTRNADLEKNDNKHLFPPLFKINGN 1618

Qy         211 LLEVISPSPHFYP------DFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYV 264
               :    | | |||        :::|  |    : :: | :  |           | :  :
Db        1619 DMNYPYPYPWFYPQQKQEDSSNQIRHLFEQQLQFMQQRYENELTELRRQRALLEKKLDQI 1678

Qy         265 QLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLGFIGKMFKSLDLSLIVEFILMFYR 324
               |||  : || |                  :| :   : :| : | |    : :  |
Db        1679 QLESQLSAKKN------------------DFEK---VEQMMQKL-LEKTEQKLSAF-- 1712

Qy         325 DKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRF----KPSLFQHVGTHSSLAGKIQ 380
               |: |: | : |      | :| |:| | :|  | :|     | :| | | |   :   :|
Db        1713 DQKINALAEQI------NTQK-AEHADSEKQQLLLRIEQLEKQNLAQAVQTPQPVQPVVQ 1765
```

Search completed: July 20, 2004, 10:12:38
Job time : 24 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 20, 2004, 10:07:10 ; Search time 13 Seconds
(without alignments)
2194.957 Million cell updates/sec

Title: US-10-033-245-24
Perfect score: 2868
Sequence: 1 MRLRNGTFLTLLLFCLCAFL.....IQTDSPVWVILSEIFLKKAD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	136	4.7	740	1	BBAP_HUMAN	Q8tdb6 homo sapien
2	123	4.3	1553	1	OSP_DROME	Q27421 drosophila
3	119.5	4.2	1290	1	RA50_SCHPO	Q9utj8 schizosacch
4	113.5	4.0	1755	1	PEPL_MOUSE	Q9r269 mus musculu
5	108	3.8	274	1	POSC_MOUSE	Q9z2y8 mus musculu
6	108	3.8	838	1	CYAA_PASMU	Q05766 pasteurella
7	107.5	3.7	970	1	T3RE_BPP1	P08764 bacteriopha
8	107	3.7	658	1	KPC1_LYTPI	Q25378 lytechinus
9	105.5	3.7	1165	1	SYV_AQUAE	O67411 aquifex aeo
10	105.5	3.7	6486	1	TYCC_BREPA	O30409 b tyrocidin
11	105	3.7	2324	1	COAC_CHICK	P11029 gallus gall
12	105	3.7	3911	1	AKA9_HUMAN	Q99996 h a-kinase
13	104.5	3.6	3770	1	ACVS_EMENI	P27742 emericella
14	104	3.6	1039	1	IF2_ANASP	Q8yqj1 anabaena sp
15	104	3.6	1818	1	HMW2_MYCPN	P75471 mycoplasma
16	104	3.6	2280	1	YCF2_OENHO	Q9mef2 oenothera h
17	103	3.6	1247	1	KPBA_DROME	Q9w391 drosophila
18	102.5	3.6	575	1	EZRA_STRPN	Q97rk0 streptococc
19	102.5	3.6	575	1	EZRA_STRR6	Q8dqe5 streptococc
20	101.5	3.5	412	1	GAS7_HUMAN	O60861 homo sapien
21	101.5	3.5	916	1	TOP1_ARATH	P30181 arabidopsis
22	101	3.5	406	1	OLFL_HUMAN	Q9nrn5 homo sapien
23	101	3.5	547	1	CITA_KLEPN	P52687 klebsiella
24	99.5	3.5	4540	1	DYHC_PARTE	Q27171 paramecium
25	99	3.5	1839	1	POLR_EPMV	P20126 eggplant mo
26	98.5	3.4	1006	1	RAT1_YEAST	Q02792 saccharomyc
27	97.5	3.4	1487	1	MDS3_YEAST	P53094 saccharomyc
28	97	3.4	1698	1	CUL7_HUMAN	Q14999 homo sapien
29	96.5	3.4	868	1	PD6I_HUMAN	Q8wum4 homo sapien
30	96	3.3	422	1	GAS7_RAT	O55148 rattus norv
31	96	3.3	865	1	SYV_THEMA	Q9x2d7 thermotoga
32	96	3.3	2663	1	CENE_HUMAN	Q02224 homo sapien
33	95.5	3.3	638	1	LIK2_RAT	P53670 rattus norv
34	95.5	3.3	869	1	PD6I_MOUSE	Q9wu78 mus musculu
35	95.5	3.3	1232	1	Y908_METJA	Q58318 methanococc
36	95	3.3	543	1	SYFB_THEAC	P57694 thermoplasm
37	95	3.3	562	1	EZRA_BACSU	O34894 bacillus su
38	95	3.3	914	1	ITH1_MESAU	P97278 mesocricetu
39	95	3.3	2345	1	COA1_RAT	P11497 rattus norv
40	94.5	3.3	4910	1	MDN1_YEAST	Q12019 saccharomyc
41	94	3.3	487	1	5HT3_MOUSE	P23979 mus musculu
42	94	3.3	591	1	SYD_PSESM	Q87y31 pseudomonas
43	94	3.3	757	1	GSH1_PASMU	Q9cm00 pasteurella
44	93.5	3.3	417	1	AGP_PRORE	Q52309 providencia
45	93.5	3.3	448	1	T2EA_SCHPO	Q9p3w1 schizosacch

ALIGNMENTS

RESULT 1
BBAP_HUMAN
ID BBAP_HUMAN STANDARD; PRT; 740 AA.
AC Q8TDB6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE B-lymphoma- and BAL-associated protein (Rhysin 2) (Rhysin2).
GN BBAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., IN VITRO UBIQUITIN LIGASE ACTIVITY, AND
RP INTERACTION WITH BAL AND DTX1.
RX MEDLINE=22679154; PubMed=12670957;
RA Takeyama K., Aguiar R.C.T., Gu L., He C., Freeman G.J., Kutok J.L.,
RA Aster J.C., Shipp M.A.;
RT "The BAL-binding protein BBAP and related Deltex family members
RT exhibit ubiquitin-protein isopeptide ligase activity.";
RL J. Biol. Chem. 278:21930-21937(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Roberts R.C., Kendrick-Jones J., Jensen O.N.;
RT "Rhysin2 is a novel protein identified by mass spectrometry found in a
RT myosin VI-containing complex isolated by immunoprecipitation.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Functions as an ubiquitin ligase protein in vitro.
CC -!- SUBUNIT: Homodimer and heterodimer. Can heterodimerize with DTX1,
CC enhancing its ubiquitin ligase activity in vitro. Interacts with
CC BAL. Found in a complex with MYO6.

```
CC   -!- SIMILARITY: Belongs to the Deltex family.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF484416; AAL90859.1; -.
DR   EMBL; BC042191; AAH42191.1; -.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
FT   ZN_FING     561    600       RING-TYPE.
SQ   SEQUENCE   740 AA;  83554 MW;  C413BB744CEE6223 CRC64;

  Query Match             4.7%;  Score 136;  DB 1;  Length 740;
  Best Local Similarity   22.3%;  Pred. No. 0.018;
  Matches  119;  Conservative   72;  Mismatches  159;  Indels  184;  Gaps   28;

Qy         28 LSGQKGDVVDVYQREFLALRDRLHAAEQESLKRSKELNLVLDE-------------IKRA 74
              :| |: :    :: ||         ||::  ||:  |  :::||              ||:
Db         43 VSTQEHEAPGTFRVEFSE-----RAAKERVLKKG-EHQILVDEKPVPIFLVPTENSIKK- 95

Qy         75 VSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLP------------TVFHHLP-H 121
               : |  :              ||:     |    |:|  |:|             ||   | :
Db         96 -NTRPQI---------SSLTQSQAETPSGDMHQHEGHIPNAVDSCLQKIFLTVTADLNCN 145

Qy        122 LLAKE-----SSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISE----LSP 172
              | :||     ::| |::|  :|  |:  | |        :| :|::    | ||     ||
Db        146 LFSKEQRAYITTLCPSIRKMEGHDGIEKVCGDFQDIERIHQFLSEQP--LESEQKQQFSP 203

Qy        173 ---------QEKEDSVIVVLIAETDSQYTS-------AVTENIKALFPTEIHS-----GL 211
                      |:: || |      || ::  |          |  | : | :|:|     |:
Db        204 SMTERKPLSQQERDSCISPSEPETKAEQKSNYFEVPLPYFEYFKYICPDKINSIEKRFGV 263

Qy        212 -LEVISPSPHFY---------PDFSRLRESF------------------GDPKERVRWRT 243
               :|:   ||:             |    ||||                | |:  ::
Db        264 NIEIQESSPNMVCLDFTSSRSGDLEAARESFASEFQKNTEPLKQECVSLADSKQANKF-- 321

Qy        244 KQNLDYCFLMMYAQSKG--IYYVQLEDDIVAKPNYLS------TMKNFA-------LQQP 288
              || |:: |  :  : ||  :  :  :||| |    :|       :| ||        ::
Db        322 KQELNHQFTKLLIKEKGGELTLLGTQDDISAAKQKISEAFVKIPVKLFAANYMMNVIEVD 381

Qy        289 SEDWMILEFSQLGFIGKMFKSLDL-SLIVE-----FILMFYRDKPID------------- 329
              |  : :||   |  | :: |  |: | : |       ||   :|: :|
Db        382 SAHYKLLETELLQEISEIEKRYDICSKVSEKGQKTCILFESKDRQVDLSVHAYASFIDAF 441

Qy        330 -----WLLDHILWVKVCNPEKDAKHCDRQKANLRIRFKPSLF------------------ 366
                  |:  :| :|    |:  |  :   : | |     |
Db        442 QHASCQLMREVLLLKSLGKERKHLHQTKFADDFRKRHPNVHFVLNQESMTLTGLPNHLAK 501

Qy        367 --QHV---GTHSSLAGKIQKLK---------DKDFGKQALRKEHVNPPAEVSTS 406
                |:|   |  ||||||  |||          | |  | |      :|| : | |
Db        502 AKQYVLKGGGMSSLAGK--KLKEGHETPMDIDSDDSKAA------SPPLKGSVS 547


RESULT 2
OSP_DROME
ID   OSP_DROME      STANDARD;      PRT;  1553 AA.
AC   Q27421; Q24231; Q8MSX4; Q9VJS4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Outspread protein.
GN   OSP OR BG:DS01486.5 OR BG:DS01486.7 OR BG:DS07721.2 OR CG3479.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=Berkeley;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
```

```
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Berkeley; TISSUE=Embryo;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.W.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R.A., Gonzalez M., Guarin H., Kronmiller B., Li P.W., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J.M., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K.H., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.E.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 117-271 FROM N.A.
RX   MEDLINE=83271489; PubMed=6410283;
RA   Kreitman M.;
RT   "Nucleotide polymorphism at the alcohol dehydrogenase locus of
RT   Drosophila melanogaster.";
RL   Nature 304:412-417(1983).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q27421-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q27421-2; Sequence=VSP_004063, VSP_004064, VSP_004065;
CC   -!- SIMILARITY: Contains 2 PH domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003410; AAF44880.1; -.
DR   EMBL; AE003410; AAF44881.1; -.
DR   EMBL; AE003644; AAF53402.3; -.
DR   EMBL; AY118512; AAM49881.1; -.
DR   EMBL; M17837; AAA70211.1; ALT_INIT.
DR   EMBL; M19547; AAA70209.1; ALT_INIT.
DR   EMBL; Z00030; CAA77329.1; ALT_INIT.
DR   FlyBase; FBgn0003016; osp.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00169; PH; 2.
DR   SMART; SM00233; PH; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
KW   Repeat; Alternative splicing.
FT   DOMAIN       71    177       PH 1.
FT   DOMAIN      493    592       PH 2.
FT   VARSPLIC      1    462       Missing (in isoform Short).
FT                                /FTId=VSP_004063.
FT   VARSPLIC   1373   1373       P -> V (in isoform Short).
FT                                /FTId=VSP_004064.
FT   VARSPLIC   1374   1553       Missing (in isoform Short).
FT                                /FTId=VSP_004065.
SQ   SEQUENCE   1553 AA;  174107 MW;  861104687ECB1DE1 CRC64;

  Query Match             4.3%;  Score 123;  DB 1;  Length 1553;
  Best Local Similarity   21.5%;  Pred. No. 0.49;
  Matches  109;  Conservative   69;  Mismatches  200;  Indels  130;  Gaps   23;

Qy         40 QREFLALRDRLHAAEQE---SLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTED 96
              :|:  ||: :|  :|:    |||  ::| | | |::|   ||:     :|       :|
Db        931 ERQVRALKQKLAKSERRRSLSLKGKEQLELKLSELQRETVERK-----EGTPPESSSSES 985

Qy         97 PRLKPWNGSHRHVLHLPTVFHHLPHLL--AKE------SSLQPAVRVGQ-GRTGVSVVMG 147
                | |       ||    | | |:|  :||      : ||  :| ||  |  || :
Db        986 SSQSPLNA------HLLQRLHSLEHVLLGSKERLEQSLTQLQ-QIRAGQRTRRSVSPMND 1038

Qy        148 IPSVRREVHSYLTDT---------------------LHSLISELSPQEKEDSVIVVLIA 185
                 |::   | :|                      |   : :||  :::
Db       1039 RKDGLRQLERALAETCVMVSEQMELTCLQDSCHKCCDLRQRVEKLSALQQQ--------T 1090

Qy        186 ETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPH---FYPDFSRLRESFGDPKERVRWR 242
              ||| | :  : |  :    |::   | :  |              | |  |  :|| | |

Db       1091 ETDLQRSEQLLEQRE----TDLAQALEKCASQEQEQELLLQQRQELSEELGRQQERCR-R 1145

Qy        243 TKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLGF 302
              :: |:   |:      |     |||      :  | :   | | :|         |  ||
Db       1146 LEKRLE---LLEREHGK-----QLE---CLREVYHTEHANAADEQSFRKRYQTEIEQLRT 1194

Qy        303 I-GKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLR--- 358
              : |   ::: |          ::   :|    | : ::    ||:    :  :| |
Db       1195 LCEKGLSAMETS---------HKRLTMDLEQKHKMEIERLLAEKETALAEETQATLAALD 1245

Qy        359 --------------IRFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVS 404
                           |||    : |     : |   |||::| |:  :          |:
Db       1246 AMRKAHQSEVQREVARFKQEFLRQVQRGEQMRGDGAKLKEEDLGELRM---------EIL 1296

Qy        405 TSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLF 464
              : |    :| | | |    | :        |:  | | |    ||:    |
Db       1297 AFSEKYSIKCVENAALEEKLHMANSKLR----HFQQMQQLELRNKQFRA----H------ 1342

Qy        465 NTSVEVLPFDNPQSDKEALQEGRTATLR 492
                   |  |:| :|   :| | |:  |
Db       1343 ------LASDDPSNDVHFVQ-GLTSDAR 1363


RESULT 3
RA50_SCHPO
ID   RA50_SCHPO     STANDARD;      PRT;  1290 AA.
AC   Q9UTJ8; Q9P3T5;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   DNA repair protein rad50.
GN   RAD50 OR SPAC1556.01C OR SPAP4C9.01C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND INTERACTION WITH RAD21 COHESIN
RP   COMPLEX.
RX   MEDLINE=21583333; PubMed=11726502;
RA   Hartsuiker E., Vaessen E., Carr A.M., Kohli J.;
RT   "Fission yeast Rad50 stimulates sister chromatid recombination and
RT   links cohesion with repair.";
RL   EMBO J. 20:6660-6671(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
```

```
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   SUBUNITS, AND SUBCELLULAR LOCATION.
RX   MEDLINE=22825616; PubMed=12944482;
RA   Chahwan C., Nakamura T.M., Sivakumar S., Russell P., Rhind N.;
RT   "The fission yeast Rad32 (Mre11)-Rad50-Nbs1 complex is required for
RT   the S-phase DNA damage checkpoint.";
RL   Mol. Cell. Biol. 23:6564-6573(2003).
CC   -!- FUNCTION: Involved in DNA double-strand break (DSB) repair.
CC       Involved in mating type switching and has a role in choosing the
CC       sister chromatid for recombinational repair. Also has a role in
CC       telomere length maintenance.
CC   -!- SUBUNIT: Interacts with the rad21 cohesin complex. Forms a
CC       multisubunit endonuclease complex, MRN, together with nbs1 and
CC       rad32.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL132984; CAB61212.1; -.
DR   EMBL; AL360094; CAB96041.1; -.
DR   PIR; T50080; T50080.
DR   GeneDB_SPombe; SPAC1556.01c; -.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR004584; Rad50.
DR   TIGRFAMs; TIGR00606; rad50; 1.
KW   DNA repair; Hydrolase; ATP-binding; Coiled coil; Nuclear protein;
KW   Telomere; Meiosis.
FT   NP_BIND      34     41       ATP (POTENTIAL).
FT   DOMAIN      295    365       COILED COIL (POTENTIAL).
FT   DOMAIN      727    809       COILED COIL (POTENTIAL).
FT   DOMAIN      814    902       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1290 AA;  149562 MW;  47BD2211E619D694 CRC64;

  Query Match             4.2%;  Score 119.5;  DB 1;  Length 1290;
  Best Local Similarity   19.3%;  Pred. No. 0.69;
  Matches  100;  Conservative   84;  Mismatches  180;  Indels  155;  Gaps   22;

Qy         34 DVVDVYQREFLALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQA--LRDGDG-NRTW 90
              ::|| | : :  :| :|   |:    :|  |   |  :| : ||  :  |:| | |
Db        512 EIVDSYHK-YSGVRTKLQVFEENKTNKSAILANQLMTLKSSFSEVMSYELKDDDNYNEEL 570

Qy         91 GRLTEDPRLKPWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPS 150
               :| || | |                     | ::   : ::|  : |  : : : : |
Db        571 DKLVEDVRKK---------------------LQEKEEAESSLRSVRERLEIRISLSVQS 608

Qy        151 VR----------REVHSYLTDTLHSLISELSPQEKE-----------------DSVIVV 182
              :           : : || : |  |:|||:   | |                 :  | :
Db        609 INDLTENKKIKTKTLKSY-SGTFASMISEIKALESEIEENRKTLHSLQFGSTFYEKAIEI 667

Qy        183 LIAETDSQYTS-AVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVRW 241
               : :   |    :: :  : ||    || :  : | |   |      | ::|
Db        668 CVDQHACQLCQRSLDKEEEKLFVEHCHSMIDVIPSKSAEVYSHLETLTKTF--------- 718

Qy        242 RTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLG 301
                 :||        :::| |:     |:|      ||  |                ::|
Db        719 ---KNL--------SEAKPIF-----DEIELLDKRLSETK---------------TELS 746

Qy        302 FIGKMFKSLDL---SLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLR 358
               :    : ||:    :  |   :: |   :: |   :|   : | |:: :  ||:   ||
Db        747 DLQGDLQGLDIRKDEIQSELDTLYLRRANLEKL--QLLVKDISNLEEEIRTIDRETEVLR 804

Qy        359 IRFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKA 418
              |    |:  |      :  : :|| :|   :  |||:                  :|:
Db        805 IELPSSIAHH--NLDEIYAEREKLLEK---RGYLRKQ------------------IERT 840

Qy        419 YLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTS----------V 468
                | |    :|      | :      | |:| :  |:   :|  |  :  :|          :
Db        841 KLEET---SFKKKIDDAVLANNEQKLKLTKLNFQVNELEQLEKDINKSSEDCDLQKKKLL 897

Qy        469 EV-------LPFDNP-QSDKEAL--------QEGRTATL 491
              ||        || |  :|: | |         |: ||  |
Db        898 EVSSKQGSQAPFLNELESEYEKLEADIQEMAQKSRTEIL 936


RESULT 4
PEPL_MOUSE
ID   PEPL_MOUSE     STANDARD;      PRT;  1755 AA.
AC   Q9R269; O70231; Q9CUT1; Q9JLZ7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Periplakin.
GN   PPL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   de Vries-Smits A.M.M., Waldmann V., Burgering B.M.T.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RA   Ryoo Y.W., Li K., Aho S., Cho B.H., Klement J.F., Uitto J.;
RT   "Mouse periplakin: genomic cloning and gene targeting.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1643-1755 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   SEQUENCE OF 1647-1755 FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=98190524; PubMed=9521878;
RA   Aho S., McLean W.H.I., Li K., Uitto J.;
RT   "cDNA cloning, mRNA expression, and chromosomal mapping of human and
RT   mouse periplakin genes.";
RL   Genomics 48:242-247(1998).
CC   -!- FUNCTION: Component of the cornified envelope of keratinocytes.
CC       May link the cornified envelope to desmosomes and intermediate
CC       filaments (By similarity).
CC   -!- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH EVPL (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ASSOCIATED WITH DESMOSOMES AND INTERMEDIATE
CC       FILAMENTS (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 2 plectin repeats.
```

```
CC   -!- SIMILARITY: Contains 3 spectrin repeats.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF126834; AAD20642.1; -.
DR   EMBL; AF116523; AAF29436.1; -.
DR   EMBL; AF116519; AAF29436.1; JOINED.
DR   EMBL; AF116520; AAF29436.1; JOINED.
DR   EMBL; AF116521; AAF29436.1; JOINED.
DR   EMBL; AF116522; AAF29436.1; JOINED.
DR   EMBL; AK014700; BAB29510.1; -.
DR   EMBL; AF013715; AAC40068.1; -.
DR   MGD; MGI:1194898; Ppl.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00681; Plectin; 1.
DR   Pfam; PF00435; spectrin; 5.
DR   SMART; SM00250; PLEC; 3.
KW   Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.
FT   DOMAIN       16    125       COILED COIL (POTENTIAL).
FT   DOMAIN      182    387       COILED COIL (POTENTIAL).
FT   REPEAT      214    315       SPECTRIN 1.
FT   REPEAT      321    483       SPECTRIN 2.
FT   REPEAT      503    610       SPECTRIN 3.
FT   DOMAIN      611    819       COILED COIL (POTENTIAL).
FT   DOMAIN      883   1644       COILED COIL (POTENTIAL).
FT   REPEAT     1650   1684       PLECTIN 1.
FT   REPEAT     1699   1734       PLECTIN 2.
FT   CONFLICT    166    166       E -> Q (IN REF. 2).
FT   CONFLICT    592    592       N -> T (IN REF. 2).
FT   CONFLICT    648    648       R -> S (IN REF. 2).
FT   CONFLICT    671    672       GK -> EQ (IN REF. 2).
FT   CONFLICT    689    721       QEHCPDLERQEAEVHKLNQRFNNLSQQVERRAQ ->
FT                                PRALPGPGAAGGRGAQAQPTFQQPQPAGGTE (IN REF.
FT                                2).
FT   CONFLICT    983    983       T -> A (IN REF. 2).
FT   CONFLICT   1325   1325       G -> R (IN REF. 2).
FT   CONFLICT   1344   1345       QR -> G (IN REF. 2).
SQ   SEQUENCE   1755 AA;  204003 MW;  3FEA343086E4CB8F CRC64;

  Query Match             4.0%;  Score 113.5;  DB 1;  Length 1755;
  Best Local Similarity   18.4%;  Pred. No. 3;
  Matches  107;  Conservative   94;  Mismatches  229;  Indels  151;  Gaps   20;

Qy         42 EFLALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLK- 100
              | | ||: | |  ::   |  |: |:   :    :|:  :::    |   :|  ||:|:
Db       1007 EVLQLREELEALRRQKGAREAEVLLLQQRVAALAAEKSRVQEKVTEREVVKLQNDPQLEA 1066

Qy        101 ---------PWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGI--- 148
                      |: |           |  | ||   |:    : :|:  |  |: :
Db       1067 EYRRLQEEHQREGTLREKQE-----EELSFLQAKLRRLEKERAMAEGKITVKEVLKVEKD 1121

Qy        149 PSVRREVHSYLTDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIH 208
              :| |||:  ||       ::    ::| : ::  |   :        || | :   ::
Db       1122 AAVEREVND-LTRQYEDEAAKARSGQREKTELLRKIWALEE-------ENAKVVVQEKVR 1173

Qy        209 SGLLEVISPSPHFYPDFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKG-------- 260
                 |:: | |    : : ||    : : : |   :|   |   :   :::|
Db       1174 ----EIVRPDPKAESEVANLRLELVEQERKFRGAEEQLKSYQSELEALRNRGPQVEVKEV 1229

Qy        261 ----IYYV----------QLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLGFIGKM 306
                  | |           :| ::|: |   :    :  : | ::   |: ::
Db       1230 TKEVIKYTTDPETEQELQRLREEIMDKTRLIERC-DLEIYQLKQEIQALKDTK------- 1281

Qy        307 FKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFKPSLF 366
                 :    :|: || |  |                              ::  :|||:
```

```
Db       1282 -PQVQTREVVQEILQFQED----------------------PQTKKEVESLRIQLSEEQK 1318

Qy        367 QHVGTHSSLAGKIQKLKDKDFGKQALRKEHV---------NPP---AEVS---------- 404
               : |       | : :|:| |:       ||| |          : |   |||:
Db       1319 KQVDLEGERASQEEKIKRKEEELAQQRKERVVRQEVVQYEDEPDLRAEVTAFTNSIDAEL 1378

Qy        405 -------TSLKTYQHFTLE-----KAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFR 452
                    |: ||   |    :   ||             : |    | ::|   |
Db       1379 RQIDKLHVELRRLQHRRAELERQLEELERER------------QARRAAELEVQRLQQR 1425

Qy        453 SGNIEHPE----DKLFNTSVEVLPFDNPQSDKEALQEGRTATLRYPRSPDGYLQIGSFYK 508
                 :|  |    :|: :|   ||  |  |: : ||   :    |: |
Db       1426 LAALEQEEAKTGEKVTHTQKVVLQQDPQQTREHALLRAQLEEERHRRQ------------ 1473

Qy        509 GVAEGEVDPAFGPLEALRLSIQTDSPVWVILSE-IFLKKAD 548
               : |||::|    | ||    : :    |: || : ::| |
Db       1474 -LLEGELEPLRRKLAALE---KAEIKEKVVFSESVQVEKGD 1510


RESULT 5
POSC_MOUSE
ID   POSC_MOUSE     STANDARD;      PRT;   274 AA.
AC   Q9Z2Y8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proline synthetase co-transcribed bacterial homolog protein.
GN   PROSC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99426086; PubMed=10496079;
RA   Ikegawa S., Isomura M., Koshizuka Y., Nakamura Y.;
RT   "Cloning and characterization of human and mouse PROSC (proline
RT   synthetase co-transcribed) genes.";
RL   J. Hum. Genet. 44:337-342(1999).
CC   -!- SIMILARITY: Belongs to the UPF0001 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB018567; BAA36843.1; -.
DR   HSSP; P38197; 1B54.
DR   MGD; MGI:1891207; Prosc.
DR   InterPro; IPR001608; UPF0001.
DR   Pfam; PF01168; Ala_racemase_N; 1.
DR   TIGRFAMs; TIGR00044; TIGR00044; 1.
DR   PROSITE; PS01211; UPF0001; FALSE_NEG.
SQ   SEQUENCE   274 AA;  30048 MW;  358C67207DB11113 CRC64;

  Query Match             3.8%;  Score 108;  DB 1;  Length 274;
  Best Local Similarity   21.6%;  Pred. No. 0.6;
  Matches   59;  Conservative   44;  Mismatches   80;  Indels   90;  Gaps   12;

Qy         41 REFLALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGR-------- 92
              |:  |:: || |      : ::|  ::|::                 | | ||:|
Db         33 RDLPAIQPRLVA-----VSKTKPADMVIEAY-----------GHGQRTFGENYVQELLE 75

Qy         93 -------LTEDPRLKPWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVV 145
                     |:  | :|               :| : ||                 :  |: :
Db         76 KASNPKILSSCPEIK--------------WHFIGHL--------------QKQNVNKL 105

Qy        146 MGIP--SVRREVHSY-LTDTLHSLISELSPQEKEDSVIVVLIAETDSQY------TSAVT 196
                 | :|  |:   | |  | | ::|   :  | |    :: :  :  ||::      | ||
Db        106 MAVPNLSMLETVDSVKLADKVNSSWQKKGPTEPLKVMVQINTSGEDSKHGLLPSETIAVV 165
```

```
Qy        197 ENIKALFPTEIHSGLLEV------ISPSPHFYPDFSRLR-------ESFGDPKERVRWRT 243
              |:|||  |:    ||: :      :|  |:  ||| ||        |  | | |:|
Db        166 EHIKASCPSLEFVGLMTIGSFGHDLSQGPN--PDFQRLLTLRRELCEKLGIPVEQVELSM 223

Qy        244 KQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNY 276
                ::|:     :|   |   |::   | : :|
Db        224 GMSMDF----QHAIEVGSTNVRIGSTIFGERDY 252


RESULT 6
CYAA_PASMU
ID   CYAA_PASMU     STANDARD;      PRT;   838 AA.
AC   Q05766;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl
DE   cyclase).
GN   CYA OR PM1811.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CNP1 / NTCC 10322;
RX   MEDLINE=92011391; PubMed=1917858;
RA   Mock M., Crasnier M., Duflot E., Dumay V., Danchin A.;
RT   "Structural and functional relationships between Pasteurella
RT   multocida and enterobacterial adenylate cyclases.";
RL   J. Bacteriol. 173:6265-6269(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
RN   [3]
RP   REVIEW.
RX   MEDLINE=93119764; PubMed=8418825;
RA   Danchin A.;
RT   "Phylogeny of adenylyl cyclases.";
RL   Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the adenylyl cyclase class-1 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M68901; AAA25532.1; -.
DR   EMBL; AE006218; AAK03895.1; -.
DR   PIR; A38172; A38172.
DR   InterPro; IPR000274; Adenylt_cyclse_1.
DR   Pfam; PF01295; Adenylate_cycla; 1.
DR   PROSITE; PS01092; ADENYLATE_CYCLASE_1_1; 1.
DR   PROSITE; PS01093; ADENYLATE_CYCLASE_1_2; 1.
KW   Lyase; cAMP biosynthesis; Complete proteome.
FT   DOMAIN        1    541       CATALYTIC (POTENTIAL).
FT   DOMAIN      547    838       REGULATORY (POTENTIAL).
FT   CONFLICT    469    469       D -> A (IN REF. 1).
FT   CONFLICT    659    660       TA -> PH (IN REF. 1).
SQ   SEQUENCE   838 AA;  96798 MW;  08D64CA7B0A30E62 CRC64;

  Query Match             3.8%;  Score 108;  DB 1;  Length 838;
  Best Local Similarity   21.4%;  Pred. No. 2.8;

  Matches   95;  Conservative   59;  Mismatches  161;  Indels  128;  Gaps   19;

Qy         47 RDRLHAAEQESLKRSK-----------ELNLVLDEIKRAVSERQALRDGDGNRTWGRLT 94
              || |   |:|:|:|              |:|  | : ||    | |
Db        125 RDDLSTKEKEALQRKTHLLKNWAKQFNIEINFYLMDQKRFRCFRYA------------- 170

Qy         95 EDPRLKPWNGSHRHVLHLPTVFHHLPHLLAK--------------ESSLQPAVRVGQ-- 137
               :|     || :::| |   :     | |                 || ::  || |
Db        171 -EPLTAENCGSAQYMLLLDEFYRSAIRLAGKPLLWLHLLIEQEENYESEVERLVRTQQIC 229

Qy        138 -------GRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVIVVLIAET-DS 189
                    | |          :   : |   :|  | :    |  ||| :|: ||  |
Db        230 LDDWVDFGGLG----------QLSANEYFGASLWQLYKGIDAPYK--SVIKILLLETYSS 277

Qy        190 QYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVRWRTKQN--- 246
              :| :  |  |   |  |: :| |   :|| || |  : |: :         |: || |
Db        278 EYPN--TYLIARQFKEELLTGKL---NPSHHFDPYLAMLQRA-------TRYLTKHNELK 325

Qy        247 -LDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDW-----MILEFSQ- 299
               | :          : :|     |       ||  :  :   ||:  ::|     :| | :|
Db        326 RLGFV-------RRSVYLKATEGMCWQDPNATNNWRLQHLQKLIQEWDWSDALIEELNQR 378

Qy        300 LGFIGKMPKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRI 359
               :  |  |    ||| :|::: ||:              |       |  ::
Db        379 ANWKIKQVKKAHNSLI-KFLMLSYRN-----------LVAFARKHKVNSSIMPQDISVLT 426

Qy        360 RFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAY 419
              |   : |:        | |||  |                ||   :: | |    | :: :
Db        427 RKLYTAFE------ELPGKITLL--------------NPQISLNLSEKNLLFFEVKGSK 465

Qy        420 LREDFFWAF--TPAAGDFIRFRF 440
                :| ::    ||:   |:: |:
Db        466 TFKDGWYVVNQTPSVAGFVQKRY 488


RESULT 7
T3RE_BPP1
ID   T3RE_BPP1      STANDARD;      PRT;   970 AA.
AC   P08764;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Type III restriction-modification system EcoPI enzyme res
DE   (EC 3.1.21.5).
GN   RES.
OS   Bacteriophage P1.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC   P1-like viruses.
OX   NCBI_TaxID=10678;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88245189; PubMed=2837577;
RA   Huembelin M., Suri B., Rao D.N., Hornby D.P., Eberle H., Pripfl T.,
RA   Kenel S., Bickle T.A.;
RT   "Type III DNA restriction and modification systems EcoP1 and EcoP15.
RT   Nucleotide sequence of the EcoP1 operon, the EcoP15 mod gene and some
RT   EcoP1 mod mutants.";
RL   J. Mol. Biol. 200:23-29(1988).
CC   -!- FUNCTION: THIS PROTEIN CUTS THE DNA SOME 25 BASE-PAIRS TO THE
CC       3'-END OF THE RECOGNITION SITE. IT IS ONLY REQUIRED FOR
CC       RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME.
CC   -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC       specific double-stranded fragments with terminal 5'-phosphates.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD.
CC   -!- SIMILARITY: WITH OTHER TYPE III RES PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X06287; CAA29615.1; -.
DR   PIR; S01352; S01352.
DR   REBASE; 988; EcoPI.
DR   InterPro; IPR006935; ResIII.
DR   Pfam; PF04851; ResIII; 1.
KW   Restriction system; Hydrolase; Nuclease; Endonuclease; Helicase.
SQ   SEQUENCE   970 AA;  111458 MW;  B599110154D723AA CRC64;

  Query Match             3.7%;  Score 107.5;  DB 1;  Length 970;
  Best Local Similarity   19.1%;  Pred. No. 3.8;
  Matches  105;  Conservative   76;  Mismatches  161;  Indels  207;  Gaps   28;

Qy         54 EQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLP 113
              |    :   ||| | ||  :| :  :  ||:|                  |:         |
Db        477 EINEILHDKELLLSLDNPRRFIFSKWTLREG-----------------WDN--------P 511

Qy        114 TVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYL-------------- 159
              ||  :  | :  |:      ||:|            :|  |: |:
Db        512 NVF-QICKLRSSGSTTSKLQEVGRG-----------LRLPVNEYMCRVKDRNFTLKYYV 558

Qy        160 ----TDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLL--- 212
                  | : ||: |::    :: |          |::|  : | |:| :| |: |  |
Db        559 DFTEKDFVDSLVKEVNESSFKERV--------PSKFTQELKEQIRAQYP-ELSSRALMNE 609

Qy        213 ----EVISPSPHFYPD--FSRLRESP---------------GDPKERVRWRTKQNLDYC 250
                  |:|  : :|     :|||:  :                | | | : |  :  :
Db        610 LFNDEIIDDNDNFKDSDAYSRLKSKYPAAFPIGVKPGKIKKATDGKRRTKMRVGKFSELK 669

Qy        251 FLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLGF---IGKMF 307
              |       | :    :|  | ::  :||  |:| |::          |:: |     | |::
Db        670 ELWELINQKAV----IEYKINSENEFLSIFKSFMLEETE------RFTKSGVHTRIDKIY 719

Qy        308 KSLDLSLIVEFI----------LMFYRDKPIDWLLDHI---LWVKVCNPEKDAKH---CD 351
                 |:::    :            | ||:      ||::   ::||       |  |    ||
Db        720 IHNDMAMSKSIVSDDDDFAKLNTMSYRE-----FLDNLSQTIFVK-----HDTLHKVFCD 769

Qy        352 -RQKANL-------RIRFKPSLFQHVGTHSS----------LAGKIQKLK---------- 383
               :   |:        ||   | |     ::|          ::| |   |
Db        770 IKDTINITEYLNIQTIRKIKSGFSKYLLNNSFNKFSLGYNLISGSIHPTKFTNADGKPLD 829

Qy        384 ---DKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYLRE----------------- 422
                | |   : :::   | :     : :    ||:  : :
Db        830 EVLSSDLG---VLQDNSKAPLDTYLFEEVFYDSELERRNITDREIQSVVVFSKIPKNSIK 886

Qy        423 -----------DFFWAFTPAAGDFIRFRF-------FQPLRLERFFFRSGNIEHPEDKLF 464
                        || :    | ||:: |            ||||        ||| :  ||
Db        887 IPVAGGYTYSPDFAYVVKTAEGDYLNFIIETKNVDSKDSLRLE----EKKKIEHAQ-ALF 941

Qy        465 N---TSVEV 470
              |    ||:|
Db        942 NQISQSVKV 950


RESULT 8
KPC1_LYTPI
ID   KPC1_LYTPI     STANDARD;      PRT;   658 AA.
AC   Q25378;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Protein kinase C (EC 2.7.1.-).
GN   PKC1.
OS   Lytechinus pictus (Painted sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Temnopleuroida; Toxopneustidae;
OC   Lytechinus.
OX   NCBI_TaxID=7653;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Rakow T.L., Shen S.S.;
RT   "Molecular cloning and characterization of protein kinase C from the
RT   sea urchin Lytechinus pictus.";
RL   Submitted (OCT-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This is a calcium-activated, phospholipid-dependent,
CC       serine- and threonine-specific enzyme (By similarity).
CC   -!- FUNCTION: PKC is activated by diacylglycerol which in turn
CC       phosphorylates a range of cellular proteins. PKC also serves as
CC       the receptor for phorbol esters, a class of tumor promoters (By
CC       similarity).
CC   -!- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG
CC       binding domains.
CC   -!- SIMILARITY: Contains 1 C2 domain.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. PKC
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U02967; AAA03447.1; -.
DR   HSSP; P05697; 1TBN.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR008973; C2_CaLB.
DR   InterPro; IPR002219; DAG_PE-bind.
DR   InterPro; IPR000961; Pkinase_C.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00168; C2; 1.
DR   Pfam; PF00130; DAG_PE-bind; 2.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00433; pkinase_C; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR00008; DAGPEDOMAIN.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00109; C1; 2.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00133; S_TK_X; 1.
DR   SMART; SM00220; S_TKC; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
DR   PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR   PROSITE; PS50081; DAG_PE_BIND_DOM_2; 2.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Transferase; Serine/threonine-protein kinase;
KW   Phorbol-ester binding; Zinc; Repeat.
FT   DOMAIN       28     77       PHORBOL-ESTER AND DAG BINDING 1.
FT   DOMAIN       93    142       PHORBOL-ESTER AND DAG BINDING 2.
FT   DOMAIN      164    251       C2 DOMAIN.
FT   DOMAIN      325    583       PROTEIN KINASE.
FT   NP_BIND     331    339       ATP (BY SIMILARITY).
FT   BINDING     354    354       ATP (BY SIMILARITY).
FT   ACT_SITE    449    449       BY SIMILARITY.
SQ   SEQUENCE   658 AA;  74871 MW;  74B5A27A49C835A2 CRC64;

  Query Match             3.7%;  Score 107;  DB 1;  Length 658;
  Best Local Similarity   19.6%;  Pred. No. 2.4;
  Matches  108;  Conservative   78;  Mismatches  138;  Indels  226;  Gaps   31;

Qy         63 ELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLPHL 122
                 :| |: |: ||   ::      |  | |||  : |  |:  : : |           |
Db        188 KLKLIPDQ-KRETKKKTRTIKGSLNPTWGE-SFDFNLEDTDRNRR-------------L 231

Qy        123 LAKESSLQPAVR---VGQGRTGVSVVM--GIPSVRREVHSYLTDTLHSLISELSPQEKED 177
                  | :      | |   :|    |:| :|  |:            |  : |:     ||: :
```

```
Db        232 LVEVWDWDRATRNDFMGALSFGISELMKAGV------------DAWYKLLG----QEEGE 275

Qy        178 SVIVVLIAETDSQYTSAVTENIKAL-FPTEIH---------SGLLEVISPSPHFYPDFSR 227

Db        276 YYNVPAIAETES--IDELTSNIKKLPMPTQEHVKPQNSNSMSG-MGVVRAS-----DFNF 327

Qy        228 L----RESFGDPKERVRWRTKQNLDYCFLMMYAQSKG---IYYVQL--------EDDI-- 270

Db        328 LSVLGKGSFGK-----------------VMLAEKKGTDELYAIKILKKDVIIQDDDVEC 369

Qy        271 ----------VAKPNYLSTMKNFALQQPSEDWMILEF-------SQLGFIGK------MF 307

Db        370 TMTEKRVLGLPSKPAFLTALHS-CFQTMDRLFFVMEFVNGGDLMFQIQKVGKFREPHAVF 428

Qy        308 KSLDLSLIVEFIL---MFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFKPS 364

Db        429 YAAEIAVGLFYLHSQGVIYRDLKLD------------NVLVDAE---------------- 460

Qy        365 LFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVN----------PPAEVSTSLKTYQHFT 414

Db        461 --------------GHIKIADFG---MCKEHMNEGDTTRTFCGTPDYIAPEIVAYQ--- 499

Qy        415 LEKAYLREDFFWAF-----------TPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKL 463

Db        500 ---PYGKAVDWWAFGVLLYEMLAGQPPFDGE----------------------DEDEL 532

Qy        464 FNTSVEVLPFDNPQSDKEALQEGRTATLRYP-----RSPDGYLQI--GSFYKGV-----A 511

Db        533 FQSIMEHVPSYPKSMSRESVTMCKGFLTKHPGKRLGSGPTGEQDIREHQFFRRIDWEKLA 592

Qy        512 EGEVDPAFGP 521

Db        593 NREIQPPFVP 602


RESULT 9
SYV_AQUAE
ID   SYV_AQUAE      STANDARD;      PRT;  1165 AA.
AC   O67411;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
GN   VALS OR AQ_1413.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC       + L-valyl-tRNA(Val).
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000739; AAC07375.1; -.
DR   PIR; A70423; A70423.
DR   HSSP; P96142; 1GAX.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         43     53       "HIGH" REGION.
FT   SITE        800    804       "KMSKS" REGION.
FT   BINDING     803    803       ATP (BY SIMILARITY).
SQ   SEQUENCE   1165 AA;  137150 MW;  E612D4F28EF4A237 CRC64;

  Query Match             3.7%;  Score 105.5;  DB 1;  Length 1165;
  Best Local Similarity   19.1%;  Pred. No. 7;
  Matches   66;  Conservative   58;  Mismatches  103;  Indels  119;  Gaps   17;

Qy        174 EKEDSVIVVLIAETDSQYTSAV----------TENIKALFPTE--------IHSGLLEV 214

Db        719 EQERDVL-------DTWFSSALWPFGVFGWPESTEDLKNLYPTDLLVTGFDIIFFWVARM 771

Qy        215 ISPSPHFYPDF--------SRLRESFG----------DPKERVRWRTKQNLDYCFLMMY 255

Db        772 IMMGTHFMKDIPFYDVYVHALVRDKYGRKMSKTIGNVIDPLDIIERYGADALRFTLAILT 831

Qy        256 AQSKGI---------------------YYVQLE--DDIVAKPNYLSTMKNFALQQPSEDW 292

Db        832 VQGRDIKLAEEKFEGYKHFANKIWNAARYVLMNTPEDFIARIPYMAPLK------PEDKW 885

Qy        293 MILEFSQLG-FIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHI---LWVKV--CNPEKD 346

Db        886 IITKLNETAEEVNKALENYQYSQAAHAIYEFFWSDYCDWYIEFTKERIYKKAPEDNEEEK 945

Qy        347 AK----------HCDRQKANLRI--RFKPSLFQHV--------GTHSSLA-------- 376

Db        946 AKVENERTTALYTLHYVLEKA-LRILHPFMPYITEELWHKLPNAEGESISLAEFPQKNED 1004

Qy        377 -------GKIQKLKDKDFGKQALRKE-HVNPPAEVSTSLKTYQHFT 414

Db       1005 EIYEEDKQKVERLKEIISAIRAIRSDLQIKPSEKIKASFKTESEFS 1050


RESULT 10
TYCC_BREPA
ID   TYCC_BREPA     STANDARD;      PRT;  6486 AA.
AC   O30409;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tyrocidine synthetase III [Includes: ATP-dependent asparagine
DE   adenylase (AsnA) (Asparagine activase); ATP-dependent glutamine
DE   adenylase (GlnA) (Glutamine activase); ATP-dependent tyrosine
DE   adenylase (TyrA) (Tyrosine activase); ATP-dependent valine adenylase
DE   (ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA)
DE   (Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine
DE   activase)].
GN   TYCC.
OS   Brevibacillus parabrevis.
OC   Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX   NCBI_TaxID=54914;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 8185 / IAM 1031 / IFO 3331 / NCDO 717 / NCIB 8598;
RX   MEDLINE=98012987; PubMed=9352938;
RA   Mootz H.D., Marahiel M.A.;
RT   "The tyrocidine biosynthesis operon of Bacillus brevis: complete
RT   nucleotide sequence and biochemical characterization of functional
RT   internal adenylation domains.";
RL   J. Bacteriol. 179:6843-6850(1997).
CC   -!- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
```

```
CC       GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
CC       PEPTIDE PRODUCT.
CC   -!- COFACTOR: Contains 6 covalently bound phosphopantetheines (By
CC       similarity).
CC   -!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC   -!- SUBUNIT: LARGE MULTIENZYME COMPLEX OF TYCA, TYCB AND TYCC.
CC   -!- DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
CC       THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC       INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC       FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC       ADENYLATION, THIOLATION, CONDENSATION (NOT FOR THE INITIATION
CC       MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC       (OPTIONAL).
CC   -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC       DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC       ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC       RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC       CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   -!- SIMILARITY: Contains 6 acyl carrier domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF004835; AAC45930.1; -.
DR   PDB; 1DNY; 17-MAY-00.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR000379; Ser_estrs.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00501; AMP-binding; 6.
DR   Pfam; PF00668; Condensation; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR   PROSITE; PS00455; AMP_BINDING; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine;
KW   Multifunctional enzyme; Repeat; 3D-structure.
FT   REPEAT      466   1038       DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT   REPEAT     1521   2070       DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT   REPEAT     2536   3113       DOMAIN 3 (TYROSINE-ACTIVATING).
FT   REPEAT     3590   4149       DOMAIN 4 (VALINE-ACTIVATING).
FT   REPEAT     4606   5203       DOMAIN 5 (ORNITHINE-ACTIVATING).
FT   REPEAT     5658   6245       DOMAIN 6 (LEUCINE-ACTIVATING).
FT   DOMAIN      970   1037       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2007   2074       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3045   3112       ACYL CARRIER (ACP) 3.
FT   DOMAIN     4080   4147       ACYL CARRIER (ACP) 4.
FT   DOMAIN     5124   5191       ACYL CARRIER (ACP) 5.
FT   DOMAIN     6167   6234       ACYL CARRIER (ACP) 6.
FT   BINDING    1000   1000       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    2037   2037       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3075   3075       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    4110   4110       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    5154   5154       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    6197   6197       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   6486 AA;  724011 MW;  4934900AF07DF786 CRC64;

  Query Match             3.7%;  Score 105.5;  DB 1;  Length 6486;
  Best Local Similarity   21.5%;  Pred. No. 77;
  Matches  113;  Conservative   73;  Mismatches  216;  Indels  123;  Gaps   27;

Qy         50 LHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWG------RLTED---PRLK 100
                |:| :| |  | :  :  : : | | |:|   ::    ||:        |  |   | ::
Db       4309 LYAGKQLSDLRIQYKDFAVWQTKLAQSDRFQKQEDFWTRTFAGEIPLLNLPHDYPRPSVQ 4368

Qy        101 PWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLT 160
               ::|     : | |  |||    | | ::           || :: |           |
Db       4369 SFDGD---TVALGT-GHHLLEQLRKLAA----------ETGTTLFM-----------VLL 4403

Qy        161 DTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPH 220
                 | |:|: : ||:      :::    :  : |  | |  :|   :    |    : |
Db       4404 AAYHVLLSKYAGQEE------IVVGTPIAGRSHADVERIVGMFVNTL--ALKNTAAGSLS 4455

Qy        221 FYPDFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTM 280
               |        |    | |:       :: || |  :      :  :|:  |:   | :  ||
Db       4456 F-------RAFLEDVKQNA-LHAFEHQDYPPEHL------VEKLQVRRDLSRNPLF-DTM 4500

Qy        281 KNFALQQPSE----DWMILEFSQLGFIGKMFKSLDL-----SLIVEF--ILMFYRDKPID 329
               :  | : :|    |  :  :   | | |   |||        |:|:|      :  : :|
Db       4501 FSLGLAESAEGEVADLKVSPYPVNGHIAKFDLSLDAMEKQDGLLVQFSYCTKLFAKETVD 4560

Qy        330 WLLDH---ILWVKVCNPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQKLKDKD 386
               | |   :|     :|       | : | : :  |    :|:  || || |     ||
Db       4561 RLAAHYVQLLQTITADP-------DIELARISVLSKAET-EHM-LHSFLATKTAYPTDKT 4611

Qy        387 FGKQALRKEHV-NPPAEVS----TSLKTYQHFT-----LEKAYLREDF------------ 424
               | | | :| |   | |::       |||         | :   |:
Db       4612 F--QKLFEEQVEKTPNEIAVLFGNEQLTYQELNAKANQLARVLRRKGVKPESTVGILVDR 4669

Qy        425 -------FWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNT-----SVEVLP 472
                    |   | | |:       ||  : |      :|  | ||  |     |     |
Db       4670 SLYMVIGMLAVLKAGGTFVPIDPDYPLERQAFM-----LEDSEAKLLLTLQKMNSQVAFP 4724

Qy        473 FDNPQSDKEALQEGRTATLRYPRSPD--GYLQIGSFYKGVAEGEV 515
               ::    | | : :  |  | :   |:   |:   |   |  :| |
Db       4725 YETFYLDTETVDQEETGNLEHVAQPENVAYIIYTSGTTGKPKGVV 4769


RESULT 11
COAC_CHICK
ID   COAC_CHICK     STANDARD;      PRT;  2324 AA.
AC   P11029;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Acetyl-CoA carboxylase (EC 6.4.1.2) (ACC) [Includes: Biotin
DE   carboxylase (EC 6.3.4.14)].
GN   ACAC.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=88139305; PubMed=2893793;
RA   Takai T., Yokoyama C., Wada K., Tanabe T.;
RT   "Primary structure of chicken liver acetyl-CoA carboxylase deduced
RT   from cDNA sequence.";
RL   J. Biol. Chem. 263:2651-2657(1988).
RN   [2]
RP   SEQUENCE OF 493-820 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=87106011; PubMed=2879745;
RA   Takai T., Wada K., Tanabe T.;
RT   "Primary structure of the biotin-binding site of chicken liver
RT   acetyl-CoA carboxylase.";
RL   FEBS Lett. 212:98-102(1987).
CC   -!- FUNCTION: Catalyzes the rate-limiting reaction in the biogenesis
CC       of long-chain fatty acids. This protein carries three functions:
CC       biotin carboxyl carrier protein, biotin carboxylase, and
CC       carboxyltransferase.
CC   -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
CC       + malonyl-CoA.
```

```
CC   -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)
CC       = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.
CC   -!- COFACTOR: Biotin.
CC   -!- ENZYME REGULATION: By phosphorylation.
CC   -!- PATHWAY: Long-chain fatty acid biosynthesis; first (rate-limiting)
CC       step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: PARTIAL TO CARBAMOYL PHOSPHATE SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03541; AAA48701.1; -.
DR   EMBL; X05019; CAA28675.1; -.
DR   PIR; A29924; A29924.
DR   HSSP; P24182; 1DV1.
DR   InterPro; IPR001882; Biotin_BS.
DR   InterPro; IPR005482; Biotin_carb_C.
DR   InterPro; IPR000089; Biotin_lipoyl.
DR   InterPro; IPR000022; Carboxyl_trans.
DR   InterPro; IPR005479; CPase_L_D2.
DR   InterPro; IPR005481; CPase_L_N.
DR   Pfam; PF02785; Biotin_carb_C; 1.
DR   Pfam; PF00364; biotin_lipoyl; 1.
DR   Pfam; PF01039; Carboxyl_trans; 1.
DR   Pfam; PF00289; CPSase_L_chain; 1.
DR   Pfam; PF02786; CPSase_L_D2; 1.
DR   PROSITE; PS00188; BIOTIN; 1.
DR   PROSITE; PS00866; CPSASE_1; 1.
DR   PROSITE; PS00867; CPSASE_2; 1.
KW   Fatty acid biosynthesis; Biotin; Ligase; Multifunctional enzyme;
KW   ATP-binding; Phosphorylation.
FT   NP_BIND     315    320       ATP (POTENTIAL).
FT   ACT_SITE    441    441       BY SIMILARITY.
FT   BINDING     786    786       BIOTIN.
FT   MOD_RES      78     78       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES      80     80       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES    1193   1193       PHOSPHORYLATION (BY SIMILARITY).
FT   DOMAIN     1936   1965       COENZYME A-BINDING (BY SIMILARITY).
SQ   SEQUENCE   2324 AA;  262717 MW;  3F1C541F01BBBEF6 CRC64;

  Query Match             3.7%;  Score 105;  DB 1;  Length 2324;
  Best Local Similarity   17.5%;  Pred. No. 20;
  Matches  106;  Conservative   82;  Mismatches  202;  Indels  216;  Gaps   25;

Qy         64 LNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLK---------------------- 100
              : |:||  ||     ||:  |     ||  :|:|:|
Db        190 VELILDIAKRI--PVQAVWAG-----WGHASENPKLPELLHKNGIAFMGPPSQAMWALGD 242

Qy        101 ----------------PWNGS-----------HRHVLHLPTVFHHLPHLLAKESSLQPAV 133
                             |||||             : :|::|   :   ::   :  |: |
Db        243 KIASSIVAQTAGIPTLPWNGSGLRVDWQENDLQKRILNVPQELYEKGYVKDADDGLRAAE 302

Qy        134 RVG--------QGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVIVVLIA 185
               ||        :|  |  :        |:|::   |   :|  ::  :     : |: :|
Db        303 EVGYPVMIKASEGGGGKGI--------RKVNN--ADDFPNLFRQVQAEVPGSPIFVMRLA 352

Qy        186 ETD---------SQYTSAVTENIKALFPTEI-----HSGLLEVISPSPHFYPDFSRLRES 231
              :             || :|:     :||  :       |  ::|     |       |  : :
Db        353 KQSRHLEVQILADQYGNAI-----SLFGRDCSVQRRHQKIIEEAPASIATSVVFEHMEQC 407

Qy        232 FGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLE---------DDIVAKPNYLSTMKN 282
                : | : :   ::|   :|:|   |:::|          ::|| |  :
Db        408 AVKLAKMVGYVSAGTVEY----LYSQDGSFYFLELNPRLQVEHPCTEMVADVNLPAAQLQ 463

Qy        283 FALQQPSEDWMILEFSQLGFIGKMFKSLDLSLIVEFILMF----YRDKPIDW-------- 330
               |:  |                        |  | :  :|:    : |  ||:
Db        464 IAMGIP----------------------LHRIKDIRVMYGVSPWGDGSIDFENSAHVPC 500

Qy        331 LLDHILWVKVC--NPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQKLKDKDFG 388
                 |::  ::   ||::  |       |  |   ::: :     : || : :  |  ||
Db        501 PRGHVIAARITSENPDEGFKPSSGTVQELNFRSNKNVWGYFSV--AAAGGLHEFADSQFG 558

Qy        389 K----QALRKEHVNPPAEVSTSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPL 444
                     |:| ::        |     |     || :            :    ||
Db        559 HCFSWGENREEAISNMVVALKELSIRGDFRTTVEYLIK-----------LLETESFQQN 606

Qy        445 R-----LERFFFRSGNIEHPE-------------DKLFNTSV----------EVLPFDNP 476
              |     |:|         | |:             |  |  ||          :|||
Db        607 RIDTGWLDRLIAEKVQAERPDTMLGVVCGALHVADVSFRNSVSNFLHSLERGQVLPAHTL 666

Qy        477 QS--DKEALQEGRTATLRYPR-SPDGYLQI----------------------GSFYKGV 510
               :  | | : |||    |:  | ||: |: |                      || |
Db        667 LNTVDVELIYEGRKYVLKVTRQSPNSYVVIMNSSCVEVDVHRLSDGGLLLSYDGSSYTTY 726

Qy        511 AEGEVD 516
               : |||
Db        727 MKEEVD 732


RESULT 12
AKA9_HUMAN
ID   AKA9_HUMAN     STANDARD;      PRT;  3911 AA.
AC   Q99996; O14869; O43355; O94895; Q9UQH3; Q9UQQ4; Q9Y6B8; Q9Y6Y2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE   (PRKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
DE   protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE   (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
DE   PKN-associated protein) (CG-NAP).
GN   AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 4).
RC   TISSUE=Brain;
RX   MEDLINE=98151389; PubMed=9482789;
RA   Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
RT   "Yotiao, a novel protein of neuromuscular junction and brain that
RT   interacts with specific splice variants of NMDA receptor subunit
RT   NR1.";
RL   J. Neurosci. 18:2017-2027(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX   MEDLINE=99219864; PubMed=10202149;
RA   Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Tasken K.,
RA   Jahnsen T., Oerstavik S.;
RT   "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RT   protein located in the centrosome, AKAP450.";
RL   EMBO J. 18:1858-1868(1999).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Brain;
RX   MEDLINE=99287934; PubMed=10358086;
RA   Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
RT   "Characterization of a novel giant scaffolding protein, CG-NAP, that
RT   anchors multiple signaling enzymes to centrosome and the Golgi
RT   apparatus.";
RL   J. Biol. Chem. 274:17267-17274(1999).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Kemmner W.A., Deiss S., Schwarz U.;
RT   "Cloning of Hyperion.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC   TISSUE=Gastric parietal cell;
```

```
RX   MEDLINE=99115654; PubMed=9915845;
RA   Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA   Trotter K.W., Milgram S.L., Goldenring J.R.;
RT   "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT   associated with centrosomes.";
RL   J. Biol. Chem. 274:3055-3066(1999).
RN   [6]
RP   SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC   TISSUE=Lymphoblast;
RA   Hinds K., Sutterer C., Becker M., Hawkins M.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RC   TISSUE=Lung;
RA   Milgram S.L., Goldenring J.R., Schmidt P.H.;
RT   "AKAP350: A multiply spliced family of proteins with centrosomal
RT   association.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
RN   [9]
RP   SEQUENCE OF 17-1800 FROM N.A.
RA   Wu X., Graves T., Bradshaw H.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Binds to type II regulatory subunits of protein kinase
CC       A. Scaffolding protein that assembles several protein kinases and
CC       phosphatases on centrosome and Golgi apparatus where physiological
CC       events can be regulated by phosphorylation state of protein
CC       substrates. Isoform 4/Yotiao is associated with the N-methyl-D-
CC       aspartate receptor and is specifically found in the neuromuscular
CC       junction (NMJ) as well as in neuronal synapses explaining that its
CC       role may be to organize postsynaptic specializations.
CC   -!- SUBUNIT: Interacts with the regulatory region of protein kinase N
CC       (PKN), protein phosphatase 2A (PP2A), protein phosphatase 1 (PP1)
CC       and the immature non-phosphorylated form of PKC epsilon.
CC   -!- SUBCELLULAR LOCATION: Centrosomal in many cell types and
CC       cytoplasmic in parietal cells.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC       Name=1;
CC         IsoId=Q99996-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
CC       Name=3; Synonyms=CG-NAP;
CC         IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
CC       Name=4; Synonyms=Yotiao;
CC         IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
CC       Name=5;
CC         IsoId=Q99996-5; Sequence=VSP_004108;
CC       Name=6; Synonyms=AKAP350;
CC         IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
CC   -!- TISSUE SPECIFICITY: Widely expressed. Isoform 4/Yotiao is highly
CC       expressed in skeletal muscle and in pancreas.
CC   -!- DOMAIN: RII-binding site, predicted to form an amphipathic helix,
CC       could participate in protein-protein interactions with a
CC       complementary surface on the R-subunit dimer.
CC   -!- CAUTION: Ref.6 sequence differs from that shown due to two
CC       frameshifts in positions 3782 and 3811.
CC   -!- CAUTION: Ref.9 sequence differs from that shown due to four
CC       frameshifts in positions 29, 1653, 1699 and 1735.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ131693; CAB40713.1; -.
DR   EMBL; AB019691; BAA78718.1; -.
DR   EMBL; AJ010770; CAA09361.1; -.
DR   EMBL; AF026245; AAB86384.1; -.
DR   EMBL; AF083037; AAD22767.1; -.
DR   EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR   EMBL; AF091711; AAD39719.1; -.
DR   EMBL; AB018346; BAA34523.1; -.
DR   EMBL; AC000066; AAC60380.1; ALT_FRAME.
DR   Genew; HGNC:379; AKAP9.
DR   MIM; 604001; -.
DR   GO; GO:0005813; C:centrosome; TAS.
DR   GO; GO:0005856; C:cytoskeleton; TAS.
DR   GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr. . .; TAS.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   GO; GO:0007268; P:synaptic transmission; TAS.
DR   GO; GO:0006810; P:transport; TAS.
KW   Coiled coil; Alternative splicing; Polymorphism.
FT   DOMAIN     2554   2567       PKA-RII SUBUNIT BINDING DOMAIN.
FT   DOMAIN      164    914       COILED COIL (POTENTIAL).
FT   DOMAIN      944   1022       COILED COIL (POTENTIAL).
FT   DOMAIN     1100   1185       COILED COIL (POTENTIAL).
FT   DOMAIN     1253   1280       COILED COIL (POTENTIAL).
FT   DOMAIN     1336   1392       COILED COIL (POTENTIAL).
FT   DOMAIN     1434   1459       COILED COIL (POTENTIAL).
FT   DOMAIN     1585   1659       COILED COIL (POTENTIAL).
FT   DOMAIN     1857   2455       COILED COIL (POTENTIAL).
FT   DOMAIN     2544   2561       COILED COIL (POTENTIAL).
FT   DOMAIN     2603   2776       COILED COIL (POTENTIAL).
FT   DOMAIN     3065   3092       COILED COIL (POTENTIAL).
FT   DOMAIN     3124   3470       COILED COIL (POTENTIAL).
FT   DOMAIN     3587   3689       COILED COIL (POTENTIAL).
FT   DOMAIN     3726   3730       POLY-LEU.
FT   DOMAIN      203    292       GLN-RICH.
FT   DOMAIN      321   1010       GLU-RICH.
FT   DOMAIN     1846   2772       GLU-RICH.
FT   VARSPLIC     17     28       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_004102.
FT   VARSPLIC   1637   1642       QLQEEI -> LATRRD (in isoform 4).
FT                                /FTId=VSP_004103.
FT   VARSPLIC   1643   3911       Missing (in isoform 4).
FT                                /FTId=VSP_004104.
FT   VARSPLIC   2175   2182       Missing (in isoform 3).
FT                                /FTId=VSP_004105.
FT   VARSPLIC   2175   2183       SADTFQKVE -> Q (in isoform 6).
FT                                /FTId=VSP_004106.
FT   VARSPLIC   2895   2907       VFGFYNMCFSTLC -> GSSIPELAHSDAYQTREICSS
FT                                (in isoform 2, isoform 3 and isoform 6).
FT                                /FTId=VSP_004107.
FT   VARSPLIC   2895   2948       Missing (in isoform 5).
FT                                /FTId=VSP_004108.
FT   VARSPLIC   3901   3911       STTQFHAGMRR -> ALSLTTSWQHHSARPTAPLFFEILSH
FT                                SLG (in isoform 6).
FT                                /FTId=VSP_004109.
FT   VARIANT    1347   1347       K -> KQ.
FT                                /FTId=VAR_010926.
FT   CONFLICT     76     76       E -> Q (IN REF. 3).
FT   CONFLICT    475    475       M -> I (IN REF. 3).
FT   CONFLICT    554    554       E -> G (IN REF. 3).
FT   CONFLICT    638    638       R -> S (IN REF. 3).
FT   CONFLICT    663    663       N -> S (IN REF. 3).
FT   CONFLICT    913    913       H -> N (IN REF. 3).
FT   CONFLICT    956    956       K -> N (IN REF. 3).
FT   CONFLICT    980    982       QKH -> PKP (IN REF. 1 AND 2).
FT   CONFLICT    997    997       Q -> P (IN REF. 1 AND 2).
FT   CONFLICT   1001   1001       Q -> P (IN REF. 1 AND 2).
FT   CONFLICT   1020   1020       N -> D (IN REF. 3).
FT   CONFLICT   1028   1028       V -> E (IN REF. 3).
```

```
FT   CONFLICT   1626   1626       R -> P (IN REF. 1 AND 2).
FT   CONFLICT   1703   1703       N -> T (IN REF. 3).
FT   CONFLICT   1707   1707       V -> G (IN REF. 3).
FT   CONFLICT   1802   1803       MISSING (IN REF. 5).
FT   CONFLICT   1843   1843       A -> P (IN REF. 3).

  Query Match             3.7%;  Score 105;  DB 1;  Length 3911;
  Best Local Similarity   19.6%;  Pred. No. 41;
  Matches  106;  Conservative   78;  Mismatches  145;  Indels  212;  Gaps   28;

Qy         27 ALSGQKGDVVDVYQREFLALRDRLHAAEQESLK------RSKELNLVLDEIKRAVSER-- 78
              ||    | :::   : :   |: ::|  | :::|        | |:|:: ::::||||
Db       1154 ALCSLKEELIFAQEEKIKELQ-KIHQLELQTMKTQETGDEGKPLHLLIGKLQKAVSEECS 1212

Qy         79 ---QAL-----------------RDGDGNRTWGRLTEDPRLKPWN------GSHRHVLH 111
                 | |                 | : :  :    ||| |: :         : | |
Db       1213 YFLQTLCSVLGEYYTPALKCEVNAEDKENSGDYISENEDPELQDYRYEVQDFQENMHTL- 1271

Qy        112 LPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRRE------VHSYLTD---- 161
              | |       ||  :: |      ||   |: :  |  :: :|      :|| :|:
Db       1272 LNKVTEEYNKLLVLQTRLSKI--WGQQTDGMKLEFGEENLPKEETEFLSIHSQMTNLEDI 1329

Qy        162 ----------------------TLHSLISELSPQEKE---------------------- 176
                                   | |||| |  | ||
Db       1330 DVNHKSKLSSLQDLEKTKLEEQVQELESLISSLQQQLKETEQNYEAEIHCLQKRLQAVSE 1389

Qy        177 ---------DSVIVVLIAETDSQYT----SAVTENIKALFPTEIHSGLLEVISPSPHFYP 223
                      |||:   | |:|:| |    | | :||          | :|
Db       1390 STVPPSLPVDSVV---ITESDAQRTMYPGSCVKKNI---------DGTIE---------- 1427

Qy        224 DFSRLRESFGDPKER--VRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIV---------- 271
               ||      ||  :|   |:   ||                |  |||:::
Db       1428 -FS---GEFGVKEETNIVKLLEKQ---------------YQEQLEEEVAKVIVSMSIAF 1467

Qy        272 -----------AKPNYLSTMKNFALQQPSEDW---MILEFSQLGFIGKMFKSLDLSLIVE 317
                        | |  |: :  |: |  : :   | |   |:||  : |:::|:    |
Db       1468 AQQTELSRISGGKENTASSKQAHAVCQQEQHYFNEMKLSQDQIGF--QTFETVDVKFKEE 1525

Qy        318 FILMFYRDKPIDWLL-DHILWVKVCNPE----KDAKHC---------DRQKANLRIRFKP 363
              |       ||:   | :|   : : | :     ::| |          : |  :    :
Db       1526 F-------KPLSKELGEHGKEILLSNSDPHDIPESKDCVLTISEEMFSKDKTFI---VRQ 1575

Qy        364 SLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFT--LEKAYLR 421
              |:   :   |  | :   | ::    : :|:| |    |       :|  |  | :|::|
Db       1576 SIHDEISVSSMDASRQLMLNEEQL--EDMRQELVRQYQE-------HQQATELLRQAHMR 1626

Qy        422 E 422
              :
Db       1627 Q 1627


RESULT 13
ACVS_EMENI
ID   ACVS_EMENI     STANDARD;      PRT;  3770 AA.
AC   P27742;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase
DE   (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine
DE   synthetase) (ACV synthetase) (ACVS).
GN   ACVA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=G191;
RX   MEDLINE=91286299; PubMed=2061333;
RA   Maccabe A.P., van Liempt H., Pallissa H., Unkles S.E., Riach M.B.R.,
RA   Pfeifer E., von Doehren H., Kinghorn J.R.;
RT   "Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase from
RT   Aspergillus nidulans. Molecular characterization of the acvA gene
RT   encoding the first enzyme of the penicillin biosynthetic pathway.";
RL   J. Biol. Chem. 266:12646-12654(1991).
CC   -!- FUNCTION: Each of the constituent amino acids of the tripeptide
CC       acv are activated as aminoacyl-adenylates with peptide bonds
CC       formed through the participation of amino acid thiolester
CC       intermediates.
CC   -!- CATALYTIC ACTIVITY: L-2-aminohexanedioate + L-cysteine + L-valine
CC       + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteinyl-D-valine +
CC       3 AMP + 3 diphosphate.
CC   -!- COFACTOR: Contains 3 covalently bound phosphopantetheines
CC       (Potential).
CC   -!- PATHWAY: Biosynthesis of penicillin and cephalosporin; first step.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   -!- SIMILARITY: Contains 3 acyl carrier domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54853; CAA38631.1; -.
DR   PIR; A40889; A40889.
DR   HSSP; P14687; 1AMU.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_S.
DR   InterPro; IPR000379; Ser_estrs.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW   Repeat; Phosphopantetheine.
FT   REPEAT      321    910       DOMAIN 1 (ADIPATE-ACTIVATING).
FT   REPEAT     1413   1993       DOMAIN 2 (CYSTEINE-ACTIVATING).
FT   REPEAT     2494   3078       DOMAIN 3 (VALINE-ACTIVATING).
FT   DOMAIN      850    919       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1929   2002       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3020   3087       ACYL CARRIER (ACP) 3.
FT   BINDING     882    882       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1965   1965       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3050   3050       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   3623   3623       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   3770 AA;  422448 MW;  CB66B6D232A58CB0 CRC64;

  Query Match             3.6%;  Score 104.5;  DB 1;  Length 3770;
  Best Local Similarity   19.7%;  Pred. No. 43;
  Matches  108;  Conservative   79;  Mismatches  195;  Indels  167;  Gaps   24;

Qy         39 YQREF---LALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTE 95
              | ||    : : : |       |: :| |:  :  ::   | : |
Db       1352 YARELFDEIVISELLQMVRDTLLQVAKHLDDPVRSLEYLSSAQMA--------------- 1396

Qy         96 DPRLKPWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVR-RE 154
                :|  || :       |    |   :  ||:: :|          |:||    |:  |:
Db       1397 --QLDAWNATD---AEFPDTTLHA--MFEKEAAQKP--------DKVAVVYEQRSLTYRQ 1441

Qy        155 VHSYLTDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEV 214
              ::       | | |::||  | :|:| ::: :::    :     | |:: |    |
Db       1442 LNERANRMAHQLKSDISP--KPNSIIALVVDKSEHMIAT-----ILAVWKT---GGAYVP 1491
```

```
Qy        215 ISPSPHFYPD--------------------FSRLRESFGDP----KERVRWRTKQNLDY 249
              | |    |||                    ||::|  |:     :  :  :|  |
Db       1492 IDPE---YPDDRIRYILEDTSAIAVISDACYLSRIQELAGESVRLYRSDISTQTDGNWSV 1548

Qy        250 CFLMMYAQSKGIYYVQLEDDIVAKPN------------YLSTMKNFALQQPSEDWMILEF 297
                : |  : |:         ||               :|  | | |:  ::| :|| |
Db       1549 SNPAPSSTSTDLAYIIYTSGTTGKPKGVMVEHHGVVNLQISLSKTFGLRD-TDDEVILSF 1607

Qy        298 SQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANL 357
              |   |                    |  :: : | ||  :       ||   |:::
Db       1608 SNYVF---------------------DHFVEQMTDAILNGQTLVMLNDAMRSDKER--- 1642

Qy        358 RIRFKPSLFQHVGTH--SSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTL 415
                  |:|:: |:  : |:|                          : |: :   |:
Db       1643 -------LYQYIETNRVTYLSG-------------------------TPSVISMYEFSR 1669

Qy        416 EKAYLR--EDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPF 473
                | :||  :    ||:    | ||   || | :  :          | :    |   : ||
Db       1670 FKDHLRRVDCVGEAFSQPVFDQIR-DTFQGLIINGY--------GPTEISITTHKRLYPF 1720

Qy        474 DNPQSDKEALQEGRTAT--------LRYPRSPDGYLQIGSFYKGVAEGEVD-PAFGPLEA 524
                ::||   |:   :|         | |    | | :|   :||| |  : |
Db       1721 PERRTDKSIGQQIGNSTSYVLNADMKRVPIGAVGELYLGG--EGVARGYHNRPEVTAERF 1778

Qy        525 LRLSIQTDS 533
              ||   ||||
Db       1779 LRNPFQTDS 1787
```

RESULT 14
IF2_ANASP
ID   IF2_ANASP   STANDARD;   PRT; 1039 AA.
AC   Q8YQJ1;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Translation initiation factor IF-2.
GN   INFB OR ALR3832.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
CC   -!- FUNCTION: One of the essential components for the initiation of
CC   protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC   hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC   Also involved in the hydrolysis of GTP during the formation of the
CC   70S ribosomal complex (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the IF-2 family.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AP003594; BAB75531.1; -.
DR   PIR; AI2284; AI2284.
DR   HAMAP; MF_00100; -; 1.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR000178; IF2.
DR   InterPro; IPR006847; IF2_N.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR009000; Translat_factor.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 2.
DR   Pfam; PF04760; IF2_N; 2.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   ProDom; PD186100; IF2; 1.
DR   TIGRFAMs; TIGR00487; IF-2; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS01176; IF2; 1.
KW   Initiation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   DOMAIN   536   688   G-DOMAIN.
FT   NP_BIND   542   549   GTP (BY SIMILARITY).
FT   NP_BIND   592   596   GTP (BY SIMILARITY).
FT   NP_BIND   646   649   GTP (BY SIMILARITY).
SQ   SEQUENCE   1039 AA; 111595 MW; 91E0B2B1038071C0 CRC64;

  Query Match   3.6%; Score 104; DB 1; Length 1039;
  Best Local Similarity   20.6%; Pred. No. 7.7;
  Matches 100; Conservative   78; Mismatches 158; Indels 150; Gaps   26;

```
Qy         62 KELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLPH 121
              ||| :   |:: |  | :|           ::||   :    |   |:|  | |    : |
Db        501 KELEI---EVETAEPEAEA----------RKVTEMIEV----GDLEHLLRRPPVVTIMGH 543

Qy        122 LLAKESSLQPAVR---VGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDS 178
              :   :::|  ::|   | | |          : : : :|  | :|          : ::
Db        544 VDHGKTTLLDSIRKTKVAAGEAG--------GITQHIGAYHVDIVH---------DGKEQ 586

Qy        179 VIVVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKER 238
              || |        :|:      :    |:|   :| | :                    :
Db        587 QIVFLDTPGHEAFTAMRARGARV---TDI--AVLVVAA------------------DDG 622

Qy        239 VRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFS 298
              || :| : :       :||: |:       ||   | :   |  |  ||  | :  | :
Db        623 VRPQTVEAIS------HAQAAGV-------PIVVAINKID--KEGA--QP--DRVKQELT 663

Qy        299 QLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVK-----VCNPEKDAK----- 348
              | |   : :    : : | |     | : :| ||: || |        ||:::|:
Db        664 QYGLTSEEWGGETIMVPVSAI----RGENLDTLLEMILLVAEVGELSANPDRNARGTVIE 719

Qy        349 -HCDRQKANLRIRFKPSLFQHVG---THSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVS 404
               | |: | :    :   |||     |  ||:: : | | |    |:  :  |:
Db        720 AHLDKAKGAVATLLIQNGTLHVGDILLAGSAFGKVRAMVD-DRG----RRVDIAGPS--- 771

Qy        405 TSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLF 464
                    | :|   | :       |||||                          | ::|
Db        772 --------FAVEVLGLSD------VPAAGD------------------------EPEVF 792

Qy        465 NTSVEVLPFDNPQSDKEALQ---EGR-TATLRYPRSPDGYL-QIGSFYKGVAEGEVDPAF 519
              :   |      : ::||: |    :|| | |    :: :| | ::    ||  :| |:
Db        793 DNEKEARALASDRADKQRLSRLLQGRVTLTTLSAQAQEGELKELNLILKGDVQGSVEAIV 852

Qy        520 GPLEAL 525
              | |: :
Db        853 GSLKQI 858
```

RESULT 15
HMW2_MYCPN
ID   HMW2_MYCPN   STANDARD;   PRT; 1818 AA.
AC   P75471;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytadherence high molecular weight protein 2 (Cytadherence accessory
DE   protein 2).
GN   HMW2 OR MPN310 OR MP526.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

```
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97252497; PubMed=9098066;
RA   Krause D.C., Proft T., Hedreyda C.T., Hilbert H., Plagens H.,
RA   Herrmann R.;
RT   "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT   pneumoniae cytoskeletal protein HMW2 and cytadherence.";
RL   J. Bacteriol. 179:2668-2677(1997).
CC   -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC       STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC       CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC       PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC       IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC       SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000051; AAB96174.1; -.
DR   EMBL; U59896; AAB52527.1; -.
DR   PIR; S73852; S73852.
DR   PhosSite; P75471; -.
KW   Cytadherence; Structural protein; Coiled coil; Complete proteome.
FT   DOMAIN       31    880       COILED COIL (POTENTIAL).
FT   DOMAIN      919   1607       COILED COIL (POTENTIAL).
FT   DOMAIN     1644   1755       COILED COIL (POTENTIAL).
FT   DOMAIN     1786   1817       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1818 AA;  215622 MW;  66DF4B08F0FCFBC0 CRC64;

  Query Match             3.6%;  Score 104;  DB 1;  Length 1818;
  Best Local Similarity   19.7%;  Pred. No. 17;
  Matches   71;  Conservative   55;  Mismatches  122;  Indels  112;  Gaps   15;

Qy         49 RLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRH 108
              :||  :|  : : :||  : : : | :|     |:
Db       1490 KLHLQKQSIISKGQELKEIKERVSRDISHTNKQRE------------------------ 1524

Qy        109 VLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVH---SYLTDTLHS 165
                        |  || :   ||                 :    ||::   | ||  : :
Db       1525 ---------ELNSLLHQNKLLQ--------------KNLAERERERINNKDSLLTQKIQT 1560

Qy        166 LISELSPQEKEDSVIVVL--IAETDSQYTSAVT-----------ENIKALFPT--EIHSG 210
                :||  |||  :: :|  :   : || : :|            : | |||   :|:
Db       1561 AKQKLS--EKEARILKLLEKMRAVEQQYQAEITRLKTRNADLEKNDNKHLFPPLFKINGN 1618

Qy        211 LLEVISPSPHFYP------DFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYV 264
               :    | | |||         :::|  |    : :: | :  |           | :  :
Db       1619 DMNYPYPYPWFYPQQKQEDSSNQIRHLFEQQLQFMQQRYENELTELRRQRALLEKKLDQI 1678

Qy        265 QLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQLGFIGKMFKSLDLSLIVEFILMFYR 324
              |||  : || |                     :| :   : :| : | |    : :  |
Db       1679 QLESQLSAKKN-------------------DFEK---VEQMMQKL-LEKTEQKLSAF-- 1712

Qy        325 DKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRF----KPSLFQHVGTHSSLAGKIQ 380
              |: |: | : |       | :| |:| | :|  | :|      | :| | | |   :   :|
Db       1713 DQKINALAEQI------NTQK-AEHADSEKQQLLLRIEQLEKQNLAQAVQTPQPVQPVVQ 1765
```

Search completed: July 20, 2004, 10:11:12
Job time : 16 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: July 20, 2004, 10:06:45 ; Search time 55 Seconds
(without alignments)
2815.201 Million cell updates/sec

Title: US-10-033-245-24
Perfect score: 2868
Sequence: 1 MRLRNGTFLTLLLFCLCAFL.....IQTDSPVWVILSEIFLKKAD 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2868	100.0	548	3	AAY96737	Aay96737 PRO1927,
2	2868	100.0	548	3	AAB24028	Aab24028 Human PRO
3	2868	100.0	548	5	ABG92711	Abg92711 Human sec
4	2868	100.0	548	5	ABG91363	Abg91363 Novel hum
5	2868	100.0	548	5	ABG31403	Abg31403 Human PRO
6	2868	100.0	548	6	ABU72379	Abu72379 Novel hum
7	2868	100.0	548	6	ABU80869	Abu80869 Human sec
8	2868	100.0	548	6	ABG73316	Abg73316 Human PRO
9	2868	100.0	548	6	ABU60815	Abu60815 Human sec
10	2868	100.0	548	6	ABU81238	Abu81238 Human PRO
11	2868	100.0	548	6	ABU62959	Abu62959 Human PRO
12	2868	100.0	548	7	ABO01896	Abo01896 Novel hum
13	2868	100.0	548	7	AAE38828	Aae38828 Human PRO
14	2865	99.9	548	2	AAW63559	Aaw63559 Human bet
15	2701	94.2	563	4	AAB94456	Aab94456 Human pro
16	1684	58.7	535	2	AAW63558	Aaw63558 Human bet
17	1684	58.7	535	4	AAE05189	Aae05189 Human dru
18	1672	58.3	535	3	AAY51250	Aay51250 Human IFN
19	1672	58.3	535	4	AAB48912	Aab48912 Bovine be
20	1669	58.2	535	2	AAW63557	Aaw63557 Bovine be
21	1098	38.3	578	4	ABB71754	Abb71754 Drosophil
22	1090	38.0	244	2	AAY73869	Aay73869 Human pro
23	1052	36.7	243	2	AAY73868	Aay73868 Human pro
24	605	21.1	464	5	ABB82145	Abb82145 Chicken a
25	598.5	20.9	508	4	ABB68205	Abb68205 Drosophil
26	564	19.7	478	3	AAY57603	Aay57603 Human pro
27	562	19.6	478	3	AAB42226	Aab42226 Human ORF
28	136	4.7	740	5	AAO17680	Aao17680 Human B-l
29	127	4.4	603	4	AAM00777	Aam00777 Human bon
30	123	4.3	900	4	ABB63162	Abb63162 Drosophil
31	123	4.3	1247	4	ABB59934	Abb59934 Drosophil
32	113.5	4.0	287	4	AAB88454	Aab88454 Human mem
33	109.5	3.8	1005	6	ABO14664	Abo14664 Novel hum
34	109	3.8	436	7	ADC97190	Adc97190 E. faeciu
35	109	3.8	933	5	ABP28348	Abp28348 Streptoco
36	108	3.8	578	6	ABU30855	Abu30855 Protein e
37	108	3.8	2324	2	AAR05707	Aar05707 Acetyl-Co
38	107.5	3.7	1288	3	AAG32186	Aag32186 Arabidops
39	107	3.7	406	4	AAB65902	Aab65902 Murine se
40	106	3.7	202	4	AAB65881	Aab65881 Murine TA
41	106	3.7	320	4	AAB65886	Aab65886 Murine TA
42	106	3.7	406	4	AAB65903	Aab65903 Murine se
43	106	3.7	406	4	AAB65829	Aab65829 Murine TA
44	105	3.7	406	4	AAB65901	Aab65901 Murine se
45	105	3.7	3899	6	ABR92048	Abr92048 Human cer

ALIGNMENTS

RESULT 1
AAY96737
ID AAY96737 standard; protein; 548 AA.
XX
AC AAY96737;
XX
DT 26-SEP-2000 (first entry)
XX
DE PRO1927, an N-acetylglucosaminyl transferase.
XX
KW PRO1927; N-acetylglucosaminyl transferase; secreted protein;
KW transmembrane protein; recombinant production; gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1. .23
FT /label= Signal_peptide
FT Modified-site 5. .9
FT /note= "N-glycosylation site"
FT Modified-site 6. .12
FT /note= "N-myristoylation site"
FT Modified-site 87. .91
FT /note= "N-glycosylation site"
FT Modified-site 103. .107
FT /note= "N-glycosylation site"
FT Modified-site 136. .142
FT /note= "N-myristoylation site"
FT Modified-site 370. .376
FT /note= "N-myristoylation site"
FT Modified-site 465. .469
FT /note= "N-glycosylation site"
FT Modified-site 509. .515
FT /note= "N-myristoylation site"
XX
PN WO200036102-A2.
XX
PD 22-JUN-2000.
XX
PF 01-DEC-1999; 99WO-US028634.
XX
PR 16-DEC-1998; 98US-0112851P.
PR 16-DEC-1998; 98US-0113145P.
PR 22-DEC-1998; 98US-0113511P.
PR 12-JAN-1999; 99US-0115558P.
PR 12-JAN-1999; 99US-0115565P.
PR 12-JAN-1999; 99US-0115733P.
PR 09-FEB-1999; 99US-0119341P.

PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119965P.
PR 02-JUN-1999; 99WO-US012252.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood WI;
XX
DR WPI; 2000-431586/37.
DR N-PSDB; AAA51267.
XX
PT Isolated nucleic acid molecule encodes a PRO polypeptide which is a
PT transmembrane polypeptide.
XX
PS Claim 12; Fig 18; 154pp; English.
XX
CC This is PRO3434, a novel N-acetylglucosaminyl transferase. The invention
CC concerns novel secreted and transmembrane proteins, designated PRO
CC polypeptides. The cDNA and gene sequences are useful in the recombinant
CC production of PRO polypeptides, as a hybridization probe to screen
CC libraries to isolate cDNAs with sequence identity to PRO polypeptides or
CC to map the gene encoding the PRO polypeptides and analyzing genetic
CC disorders. The cDNA/gene can also be used to produce transgenic animals
CC useful for the development and screening of therapeutically useful
CC reagents. They can also be used in gene therapy, e.g. to replace a
CC defective gene
XX
SQ Sequence 548 AA;

Query Match 100.0%; Score 2868; DB 3; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.1e-270;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548
```

RESULT 2
AAB24028
ID AAB24028 standard; protein; 548 AA.
XX
AC AAB24028;
XX
DT 25-JAN-2001 (first entry)
XX
DE Human PRO1927 protein sequence SEQ ID NO:25.
XX
KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW identification; tumourigenesis; anticancer; detection.
XX
OS Homo sapiens.
XX
PN WO200053750-A1.
XX
PD 14-SEP-2000.
XX
PF 02-DEC-1999; 99WO-US028551.
XX
PR 08-MAR-1999; 99WO-US005028.
PR 01-SEP-1999; 99WO-US020111.
PR 29-OCT-1999; 99US-0162506P.
PR 30-NOV-1999; 99WO-US028313.
PR 01-DEC-1999; 99WO-US028634.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
XX
DR WPI; 2000-594320/56.
DR N-PSDB; AAC58110.
XX
PT Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression.
XX
PS Claim 61; Fig 16; 226pp; English.
XX
CC The present invention describes an antibody that binds to a human protein
CC (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780; PRO3434;
CC PRO1927; PRO3567; PRO1295; PRO1293; PRO1303; PRO4344; PRO4354; PRO4397;
CC PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has anticancer
CC activity and can be used to diagnose tumours in mammals, by detecting
CC complex formation when the antibody is contacted with test cells.
CC Increased expression of genes encoding (I) can also be detected to
CC diagnose tumours. Agents which inhibit the activity of (I), especially
CC the antibodies, or an antisense oligonucleotide which hybridises to genes
CC encoding (I), can be used to inhibit tumour growth, preferably by
CC inducing cell death. Methods from the present invention can be used to
CC identify compounds which inhibit the biological activity of (I). AAC58019
CC to AAC58102 represent PCR primers and hybridisation probes used in
CC examples from the present invention for human PRO sequences. AAC58103 to
CC AAC58122 and AAB24021 to AAB24040 represent human PRO polynucleotide and
CC protein sequences given in the exemplification of the present invention
XX
SQ Sequence 548 AA;

Query Match 100.0%; Score 2868; DB 3; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.1e-270;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
```

```
Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548
```

RESULT 3
ABG92711
ID ABG92711 standard; protein; 548 AA.
XX
AC ABG92711;
XX
DT 18-NOV-2002 (first entry)
XX
DE Human secreted protein PRO1927.
XX
KW Human; secreted and transmembrane protein; PRO1800; PRO539; PRO982;
KW PRO1434; PRO1863; PRO1917; PRO1868; PRO3434; PRO1927;
KW inflammatory disorder; immune related disease; rheumatoid arthritis;
KW systemic lupus erythematosus; systemic sclerosis; thyroiditis;
KW autoimmune haemolytic anaemia; diabetes mellitus; infectious hepatitis;
KW psoriasis; allergic disease of the lung; graft-versus host disease;
KW tumour; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002098506-A1.
XX
PD 25-JUL-2002.
XX
PF 27-DEC-2001; 2001US-00033301.
XX
PR 04-AUG-1998; 98US-0095325P.
PR 16-DEC-1998; 98US-0112851P.
PR 16-DEC-1998; 98US-0113145P.
PR 22-DEC-1998; 98US-0113511P.
PR 12-JAN-1999; 99US-0115558P.
PR 12-JAN-1999; 99US-0115565P.
PR 12-JAN-1999; 99US-0115733P.
PR 09-FEB-1999; 99US-0119341P.
PR 10-FEB-1999; 99US-0119537P.
PR 12-FEB-1999; 99US-0119965P.
PR 02-JUN-1999; 99WO-US012252.
PR 29-OCT-1999; 99US-0162506P.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 09-DEC-1999; 99US-0170262P.
PR 11-FEB-2000; 2000WO-US003565.
PR 22-FEB-2000; 2000WO-US004414.
PR 02-MAR-2000; 2000WO-US005841.
PR 03-MAR-2000; 2000US-0187202P.
PR 30-MAR-2000; 2000WO-US008439.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 01-DEC-2000; 2000WO-US032678.
PR 25-MAY-2001; 2001US-00866034.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI Wood WI;
XX
DR WPI; 2002-690475/74.
DR N-PSDB; ABS68394.
XX
PT Novel secreted and transmembrane polypeptides and polynucleotides useful
PT for diagnosis and treatment of inflammatory disorders and immune-related
PT diseases, and identifying modulators.
XX
PS Claim 12; Fig 18; 125pp; English.
XX
CC The invention relates to an isolated polypeptide having at least 80%
CC amino acid sequence identity to secreted and transmembrane polypeptides
CC PRO1800, PRO539, PRO982, PRO1434, PRO1863, PRO1917, PRO1868, PRO3434 or
CC PRO1927 and their encoding nucleic acids. Also included are vectors, host
CC cells and antibodies against PRO polypeptides. PRO proteins are useful
CC for identifying modulators of the polypeptide. PRO1868 useful for the
CC diagnosis and treatment of inflammatory and immune related diseases
CC including systemic lupus erythematosus, rheumatoid arthritis, systemic
CC sclerosis, autoimmune haemolytic anaemia, thyroiditis, diabetes mellitus,
CC infectious hepatitis, psoriasis, allergic diseases of the lung and graft-
CC versus host disease and tumours. Pro nucleic acids are useful for
CC constructing hybridisation probes for mapping the gene that encodes that
CC PRO and for the genetic analysis of individuals with genetic disorders,
CC and for generating transgenic animals which are useful in the development
CC and screening of therapeutically useful reagents. PRO nucleic acids are
CC also useful for gene therapy, chromosome identification, and tissue
CC typing. PRO proteins are useful as molecular weight markers for protein
CC electrophoresis purposes. The anti-PRO antibodies are useful in
CC diagnostic assays for PRO, e.g. detecting its expression in specific
CC cells, tissues or serum and for affinity purification of PRO. The present
CC sequence represents a PRO protein
XX
SQ Sequence 548 AA;

```
  Query Match            100.0%;  Score 2868;  DB 5;  Length 548;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 4
ABG91363
ID   ABG91363 standard; protein; 548 AA.
XX
AC   ABG91363;
XX
DT   29-NOV-2002 (first entry)
XX
DE   Novel human secreted protein #9.
XX
KW   Human; secreted protein; transmembrane protein; gene mapping; transgenic;
KW   immunogenic.
XX
OS   Homo sapiens.
XX
PN   US2002098505-A1.
XX
PD   25-JUL-2002.
XX
PF   28-DEC-2001; 2001US-00033246.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI   Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2002-665999/71.
DR   N-PSDB; ABS67462.
XX
PT   New human secreted and transmembrane (PRO) polypeptides, useful for
PT   treating conditions requiring PRO polypeptides, for screening PRO
PT   antagonists and agonists useful as drug candidates.
XX
PS   Claim 12; Fig 18; 125pp; English.
XX
CC   The invention relates to new human secreted and transmembrane proteins
CC   (PRO) and nucleic acids of the invention. The polypeptides can be
CC   administered therapeutically, especially by expressing encoding
CC   polynucleotides, e.g. in therapeutic compositions. They can be used to
CC   screen for PRO polypeptide antagonists and agonists useful to identify
CC   drug candidates. They can also be used to produce antibodies, useful to
CC   detect PRO polypeptides (e.g. diagnostically), purify PRO polypeptides or
CC   therapeutically (e.g. as antagonists or to target and/or deliver
CC   cytotoxic agents). The polynucleotides are useful therapeutically e.g. to
CC   produce antisense sequences to inhibit polypeptide production. They can
CC   be used to produce probes and primers useful to detect or isolate
CC   sequences encoding PRO polypeptides or similar sequences e.g. variants or
CC   sequences from other species. They are also useful for gene mapping and
CC   to generate transgenic animals. ABG91355-ABG91363 represent human PRO
CC   amino acid sequences of the invention
XX
SQ   Sequence 548 AA;

  Query Match            100.0%;  Score 2868;  DB 5;  Length 548;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548
```

```
RESULT 5
ABG31403
ID   ABG31403 standard; protein; 548 AA.
XX
AC   ABG31403;
XX
DT   29-NOV-2002  (first entry)
XX
DE   Human PRO1927 polypeptide.
XX
KW   Human; secreted and transmembrane polypeptide; PRO polypeptide;
KW   T-lymphocyte proliferation; inflammatory disease; rheumatoid arthritis;
KW   inflammatory bowel disease; Sjogren's syndrome; thyroiditis;
KW   autoimmune haemolytic anaemia; diabetes mellitus; multiple sclerosis;
KW   hepatitis; contact dermatitis; allergic disease; psoriasis; virucide;
KW   immune related disease; kidney disease; antiinflammatory; antithyroid;
KW   antirheumatic; antiarthritic; immunosuppressive; antianaemic;
KW   antidiabetic; neuroprotective; hepatotropic; antiinflammatory;
KW   dermatological; antiallergic; antipsoriatic; PRO1927.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Signal_peptide
FT   Modified-site   5. .9
FT                   /note= "N-glycosylation site"
FT   Modified-site   6. .12
FT                   /note= "N-myristoylation site"
FT   Protein         24. .548
FT                   /label= Mature_PRO1927
FT   Modified-site   87. .91
FT                   /note= "N-glycosylation site"
FT   Modified-site   103. .107
FT                   /note= "N-glycosylation site"
FT   Modified-site   136. .142
FT                   /note= "N-myristoylation site"
FT   Modified-site   370. .376
FT                   /note= "N-myristoylation site"
FT   Modified-site   465. .469
FT                   /note= "N-glycosylation site"
FT   Modified-site   509. .515
FT                   /note= "N-myristoylation site"
XX
PN   US2002098507-A1.
XX
PD   25-JUL-2002.
XX
PF   27-DEC-2001; 2001US-00033326.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH') GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2002-673823/72.
DR   N-PSDB; ABS53479.
XX
PT   Novel PRO polypeptides and nucleic acids encoding the polypeptides,
PT   useful for preparing a medicament for the treatment of inflammatory and
PT   immune related disorders.
XX
PS   Claim 12; Fig 18; 125pp; English.
XX
CC   The present invention relates to the isolation of novel human secreted
CC   and transmembrane polypeptides, designated PRO polypeptides, and the
CC   polynucleotide sequences encoding them. The PRO polypeptides of the
CC   invention include PRO1800, PRO539, PRO982, PRO1434, PRO1863, PRO1917,
CC   PRO1868, PRO3434 and PRO1927. The PRO polypeptides can inhibit the
CC   stimulation of T-lymphocyte proliferation. The PRO polypeptides are
CC   useful for the diagnosis and treatment of inflammatory diseases (e.g.
CC   inflammatory bowel disease, rheumatoid arthritis, Sjogren's syndrome,
CC   autoimmune haemolytic anaemia, thyroiditis, diabetes mellitus, multiple
CC   sclerosis, hepatitis, contact dermatitis, allergic diseases and
CC   psoriasis), immune related diseases, and kidney diseases in humans. The
CC   present sequence represents human PRO1927 polypeptide
XX
SQ   Sequence 548 AA;

  Query Match            100.0%;  Score 2868;  DB 5;  Length 548;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
```

```
Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 6
ABU72379
ID   ABU72379 standard; protein; 548 AA.
XX
AC   ABU72379;
XX
DT   16-JUN-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO1927.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   antiinflammatory; dermatological; immunosuppressive; antirheumatic;
KW   antiarthritic; haemostatic; antithyroid; neuroprotective; hepatotropic;
KW   virucide; antipsoriatic; antiallergic; gene therapy; colon cancer;
KW   inflammatory bowel disease; systemic lupus erythematosus; hepatitis;
KW   rheumatoid arthritis; scleroderma; Sjogren's syndrome; thyroiditis;
KW   thrombocytopaenia; multiple sclerosis; cystic fibrosis; psoriasis;
KW   allergy; graft-versus-host disease; graft rejection.
XX
OS   Homo sapiens.
XX
PN   US2002182618-A1.
XX
PD   05-DEC-2002.
XX
PF   27-DEC-2001; 2001US-00033167.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-328610/31.
DR   N-PSDB; ACA64086.
XX
PT   New secreted and transmembrane PRO polypeptides or genes encoding them,
PT   useful for treating e.g. colon cancer, inflammatory bowel disease,
PT   Sjogren's syndrome, thrombocytopenia, thyroiditis, multiple sclerosis or
PT   graft rejection.
XX
PS   Claim 12; Fig 18; 119pp; English.
XX
CC   The invention describes an isolated secreted and transmembrane
CC   polypeptide (PRO), which scores at least 80% amino acid sequence identity
CC   when compared to: (a) a sequence comprising 278, 830, 125, 325, 437, 487,
CC   310, 1029 or 548 amino acids fully defined in the specification; (b) any
CC   of the sequences of (a), lacking its associated signal peptide; (c) an
CC   extracellular domain of (a), with or lacking its associated signal
CC   peptide. The PRO polypeptide or polynucleotide is useful as
CC   pharmaceuticals or diagnostics. These are particularly useful for
CC   treating colon cancer, inflammatory bowel disease, systemic lupus
CC   erythematosus, rheumatoid arthritis, scleroderma, Sjogren's syndrome,
CC   thrombocytopaenia, thyroiditis, multiple sclerosis, hepatitis, cystic
CC   fibrosis, psoriasis, allergies, graft-versus-host disease or graft
CC   rejection in a mammal. This is the amino acid sequence of a novel human
CC   secreted and transmembrane PRO polypeptide
XX
SQ   Sequence 548 AA;

  Query Match             100.0%;  Score 2868;  DB 6;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 7
ABU80869
ID   ABU80869 standard; protein; 548 AA.
XX
AC   ABU80869;
XX
DT   23-JUN-2003  (first entry)
XX
DE   Human secreted and transmembrane polypeptide PRO1927.
XX
KW   Human; gene therapy; inflammatory disease; Crohn's disease;
```

```
KW   inflammatory bowel disease; ulcerative colitis; tumour; cancer;
KW   colorectal cancer.
XX
OS   Homo sapiens.
XX
PN   US2002192668-A1.
XX
PD   19-DEC-2002.
XX
PF   27-DEC-2001; 2001US-00033244.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-328857/31.
DR   N-PSDB; ACA66979.
XX
PT   New secreted and transmembrane nucleic acids and polypeptides, designated
PT   as PRO, useful for treating inflammatory diseases, tumors or cancer.
XX
PS   Claim 12; Fig 18; 119pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding a PRO
CC   polypeptide. The nucleic acids and polypeptides are useful for treating
CC   inflammatory diseases such as inflammatory bowel disease, ulcerative
CC   colitis and Crohn's disease, tumours, or cancer such as colorectal
CC   cancer. The nucleic acids are useful as hybridisation probes, in
CC   chromosome and gene mapping and in generating antisense RNA or DNA. The
CC   polypeptides are useful as pharmaceuticals, diagnostics, biosensors or
CC   bioreactors. Both are useful in tissue typing. The present sequence
CC   represents the amino acid sequence of a PRO polypeptide of the invention
XX
SQ   Sequence 548 AA;

  Query Match             100.0%;  Score 2868;  DB 6;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 8
ABG73316
ID   ABG73316 standard; protein; 548 AA.
XX
AC   ABG73316;
XX
DT   30-APR-2003  (first entry)
XX
DE   Human PRO1927 polypeptide.
XX
KW   Human; secreted and transmembrane polypeptide; PRO polypeptide;
KW   inflammatory disease; immune-related disease; diabetes mellitus;
KW   rheumatoid arthritis; glomerulonephritis; multiple sclerosis;
KW   immune-mediated skin disease; contact dermatitis; graft rejection;
KW   transplantation associated disease; graft-versus-host disease;
KW   tumour diagnosis; tumour cell; antiinflammatory; immunosuppressive;
KW   cytostatic; antianaemic; antirheumatic; antiarthritic; antithyroid;
KW   antidiabetic; nephrotropic; antipsoriatic; dermatological; haemostatic;
KW   hepatotropic; virucide; neuroprotective; PRO1927.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Signal_peptide
FT   Protein         24. .548
FT                   /label= Mature_PRO1927_polypeptide
XX
PN   US2002164646-A1.
XX
PD   07-NOV-2002.
XX
PF   27-DEC-2001; 2001US-00033223.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
```

```
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-238305/23.
DR   N-PSDB; ABX11175.
XX
PT   New PRO polypeptides and nucleic acid molecules, useful in diagnosing or
PT   treating inflammatory diseases or immune-related diseases, e.g.
PT   inflammatory bowel disease, systemic lupus erythematosus or rheumatoid
PT   arthritis.
XX
PS   Claim 12; Fig 18; 119pp; English.
XX
CC   The present invention relates to the isolation of novel human secreted
CC   and transmembrane polypeptides designated PRO polypeptides (PRO1800,
CC   PRO539, PRO982, PRO1434, PRO1863, PRO1917, PRO1868, PRO3434 and PRO1927),
CC   and the polynucleotide sequences encoding them. The PRO polypeptides and
CC   polynucleotide sequences of the invention are useful in diagnosing or
CC   treating inflammatory diseases or immune-related diseases (e.g.
CC   inflammatory bowel disease, systemic lupus erythematosus, rheumatoid
CC   arthritis, Sjogren's syndrome, autoimmune haemolytic anaemia, autoimmune
CC   thrombocytopaenia, thyroiditis, diabetes mellitus, glomerulonephritis,
CC   multiple sclerosis, infectious hepatitis, immune-mediated skin diseases
CC   including psoriasis or contact dermatitis, and transplantation associated
CC   diseases including graft rejection or graft-versus-host disease). The PRO
CC   polypeptides are also useful for diagnosing tumours, and for inhibiting
CC   the growth of tumour cells. The PRO polynucleotide sequences may be used
CC   as hybridisation probes in chromosome and gene mapping, and in generating
CC   antisense RNA and DNA. They are also useful in preparing PRO
CC   polypeptides, in assays to identify other proteins or molecules involved
CC   in a binding reaction, to generate transgenic animals or knockout
CC   animals, which in turn are useful in the development and screening of
CC   therapeutically useful reagents, for chromosome identification, and
CC   tissue typing. The PRO polynucleotide sequences are also useful in gene
CC   therapy. Anti-PRO antibodies may be used in diagnostic assays for PRO
CC   polypeptides. The present sequence represents human PRO1927 polypeptide
XX
SQ   Sequence 548 AA;

  Query Match            100.0%;  Score 2868;  DB 6;  Length 548;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 9
ABU60815
ID   ABU60815 standard; protein; 548 AA.
XX
AC   ABU60815;
XX
DT   06-MAY-2003  (first entry)
XX
DE   Human secreted/transmembrane protein, #9.
XX
KW   Human; PRO; secreted; transmembrane; pharmaceutical; diagnostic;
KW   biosensor; bioreactor; therapeutic; gene therapy; tumour;
KW   inflammatory disease; immune-related disease; inflammatory bowel disease;
KW   IBD; systemic lupus erythematosus; rheumatoid arthritis; thyroiditis;
KW   diabetes mellitus; glomerulonephritis; multiple sclerosis; cirrhosis;
KW   psoriasis; graft rejection; antiinflammatory; immunosuppressive;
KW   neuroprotective; hepatotropic.
XX
OS   Homo sapiens.
XX
PN   US2002160392-A1.
XX
PD   31-OCT-2002.
XX
PF   27-DEC-2001; 2001US-00033245.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
```

```
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D, Desnoyers L, Ferrara N, Fong S, Gao W, Goddard A;
PI   Gurney AL, Pan J, Roy MA, Stewart TA, Tumas D, Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-275292/27.
DR   N-PSDB; ABX90611.
XX
PT   New isolated PRO polypeptide, e.g. PRO1800 or PRO539, useful for
PT   diagnosing, preventing and treating tumors and inflammatory or immune-
PT   related diseases, e.g. systemic lupus erythematosus, thyroiditis,
PT   diabetes or psoriasis.
XX
PS   Claim 12; Fig 18; 119pp; English.
XX
CC   The invention discloses isolated PRO secreted/transmembrane polypeptides
CC   comprising a sequence without signal peptide and the nucleic acid
CC   encoding them. The polypeptides can be used to raise antibodies that
CC   specifically bind to the PRO polypeptide, for linking a bioactive
CC   molecule to a cell expressing a PRO protein and for modulating at least
CC   one biological activity of a cell. The PRO polypeptides and the antibody
CC   are useful for diagnosing, preventing and treating tumours and
CC   inflammatory or immune-related diseases, such as inflammatory bowel
CC   disease (IBD), systemic lupus erythematosus, rheumatoid arthritis,
CC   thyroiditis, diabetes mellitus, glomerulonephritis, multiple sclerosis,
CC   cirrhosis, psoriasis or graft rejection. The proteins and the antibody
CC   may also be used in preparing medicines and medicaments for treating the
CC   above-mentioned diseases. The polynucleotide is useful in molecular
CC   biology, including uses as hybridisation probes, in chromosome and gene
CC   mapping, in generating antisense RNA and DNA, and in gene therapy. The
CC   polynucleotide may also be used in preparing PRO polypeptides by
CC   recombinant techniques, and in generating either transgenic animals or
CC   knock-out animals which, in turn, are useful in the development and
CC   screening of therapeutically useful reagents. The sequences presented in
CC   ABU60807-ABU60815 are the human PRO polynucleotides of the invention
XX
SQ   Sequence 548 AA;

  Query Match             100.0%;  Score 2868;  DB 6;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMPKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMPKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 10
ABU81238
ID   ABU81238 standard; protein; 548 AA.
XX
AC   ABU81238;
XX
DT   23-JUN-2003  (first entry)
XX
DE   Human PRO3434 polypeptide.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein; antianaemic;
KW   inflammatory disease; immune related disease; rheumatoid arthritis;
KW   juvenile chronic arthritis; scleroderma; Sjogren's syndrome; sarcoidosis;
KW   autoimmune haemolytic anaemia; thyroiditis; psoriasis; Grave's disease;
KW   diabetes mellitus; immune-mediated renal disease; glomerulonephritis;
KW   demyelinating disease; nervous system; antithyroid;
KW   hepatobiliary disease; hepatitis; primary biliary cirrhosis;
KW   fibrotic lung disease; bullous skin disease; allergic disease;
KW   pulmonary fibrosis; transplantation associated disease; haemostatic;
KW   graft rejection; graft-versus host disease; cytostatic; dermatological;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antidiabetic; nephrotropic; neuroprotective; hepatotropic; antipsoriatic;
KW   antiallergic.
XX
OS   Homo sapiens.
XX
PN   US2003032060-A1.
XX
PD   13-FEB-2003.
XX
PF   27-DEC-2001; 2001US-00032990.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
```

```
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-341961/32.
DR   N-PSDB; ACA67369.
XX
PT   Novel isolated PRO polypeptides e.g. PRO1800, PRO539 and PRO982, of N-
PT   acetylglucosaminyltransferase protein family, useful for diagnosing,
PT   treating or preventing immune disorders and inflammatory disorders.
XX
PS   Claim 12; Fig 18; 124pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The PRO
CC   polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful for preparing a medicament
CC   useful in the treatment of inflammatory and immune related diseases such
CC   as inflammatory bowel disease, systemic lupus erythematosus (SLE),
CC   rheumatoid arthritis, juvenile chronic arthritis, spondyloarthropathies,
CC   scleroderma, idiopathic inflammatory myopathies, Sjogren's syndrome,
CC   systemic vasculitis, sarcoidosis, autoimmune haemolytic anaemia,
CC   autoimmune thrombocytopaenia, thyroiditis, Grave's disease, diabetes
CC   mellitus, immune-mediated renal disease, glomerulonephritis,
CC   demyelinating diseases of the central and peripheral nervous systems such
CC   as multiple sclerosis, idiopathic polyneuropathy, hepatobiliary diseases
CC   such as infectious hepatitis, autoimmune chronic active hepatitis,
CC   primary biliary cirrhosis, granulomatous hepatitis, sclerosing
CC   cholangitis, inflammatory and fibrotic lung diseases, gluten-sensitive
CC   enteropathy, Whipple's disease, autoimmune or immune-mediated skin
CC   diseases including bullous skin diseases, erythema multiforme and contact
CC   dermatitis, psoriasis, allergic diseases of the lung such as eosinophilic
CC   pneumonias, idiopathic pulmonary fibrosis and hypersensitivity
CC   pneumonitis, and transplantation associated diseases including graft
CC   rejection and graft-versus host disease. Anti-PRO antibodies are useful
CC   in diagnostic assays for PRO, in affinity purification of PRO, and for
CC   detection of PRO in biological samples. ABU81230-ABU81238 represent the
CC   human PRO polypeptides of the invention
XX
SQ   Sequence 548 AA;

  Query Match             100.0%;  Score 2868;  DB 6;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 11
ABU62959
ID   ABU62959 standard; protein; 548 AA.
XX
AC   ABU62959;
XX
DT   14-SEP-2003  (first entry)
XX
DE   Human PRO1927 protein.
XX
KW   Human; PRO1927; immunosuppressive; antiinflammatory; gene therapy;
KW   immune disorder; inflammatory disorder; IBD; inflammatory bowel disease;
KW   glycosyltransferase; enzyme.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Signal_peptide
FT   Modified-site   5. .9
FT                   /note= "N Glycosylation site"
FT   Modified-site   6. .12
FT                   /label= N_myristoylation_site
FT   Protein         24. .548
FT                   /label= Mature_protein
FT   Modified-site   87. .91
FT                   /note= "N Glycosylation site"
FT   Modified-site   103. .107
FT                   /note= "N Glycosylation site"
FT   Modified-site   136. .142
FT                   /label= N_myristoylation_site
FT   Modified-site   370. .376
FT                   /label= N_myristoylation_site
FT   Modified-site   465. .469
FT                   /note= "N Glycosylation site"
FT   Modified-site   509. .515
FT                   /label= N_myristoylation_site
XX
PN   US2003054447-A1.
XX
PD   20-MAR-2003.
XX
PF   27-DEC-2001; 2001US-00032996.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
```

```
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-503626/47.
DR   N-PSDB; ACD26815.
XX
PT   New secreted and transmembrane PRO polypeptides and nucleic acids
PT   encoding the polypeptides, useful in gene therapy, in diagnosing and/or
PT   treating immune-related and inflammatory diseases, or in identifying
PT   chromosomes.
XX
PS   Claim 12; Fig 18; 118pp; English.
XX
CC   This invention relates to novel isolated nucleic acid sequences encoding
CC   secreted or transmembrane PRO polypeptides. Also disclosed is a method
CC   for producing a PRO polypeptide by culturing a host cell containing an
CC   expression vector comprising the full length nucleic acid sequence
CC   encoding the PRO protein. The protein and nucleic acid sequences of the
CC   invention may have immunosuppressive or antiinflammatory activities and
CC   may be used in gene therapy. Nucleic acids that encode PRO can be used to
CC   generate either transgenic animals or knock-out animals useful in
CC   developing and screening of therapeutically useful reagents. The nucleic
CC   acids may also be used in gene therapy for replacing a defective gene, in
CC   chromosome identification, as chromosome markers, or in generating probes
CC   to isolate full length PRO cDNA. The PRO polypeptides are useful as
CC   molecular markers for protein electrophoresis, and the isolated nucleic
CC   acids may be used for recombinantly expressing those markers. The PRO
CC   polypeptides and nucleic acids may also be used in tissue typing. Anti-
CC   PRO antibodies are useful in diagnostic assays for PRO, and in affinity
CC   purification of PRO from recombinant cell culture or natural sources. The
CC   proteins may also be used in the diagnosis and/or treatment of immune
CC   related diseases and inflammatory diseases (e.g. inflammatory bowel
CC   disease). The present sequence represents the human PRO1927 protein of
CC   the invention, this protein is a newly identified member of the
CC   glycosyltranferase family of proteins and may possess glycosylation
CC   activity
XX
SQ   Sequence 548 AA;

  Query Match             100.0%;  Score 2868;  DB 6;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 12
ABO01896
ID   ABO01896 standard; protein; 548 AA.
XX
AC   ABO01896;
XX
DT   07-AUG-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO1927.
XX
KW   Human; secreted and transmembrane protein; PRO; antiinflammatory;
KW   dermatological; immunosuppressive; antirheumatic; antiarthritic;
KW   antithyroid; antidiabetic; neuroprotective; hepatotropic; virucide;
KW   cytostatic; gene therapy; antisense therapy; inflammatory bowel disease;
KW   systemic lupus erythematosis; rheumatoid arthritis; systemic sclerosis;
KW   Sjogren's syndrome; autoimmune thrombocytopenia; thyroiditis;
KW   diabetes mellitus; multiple sclerosis; hepatitis; erythema multiforme;
KW   contact dermatitis; graft-versus-host-disease; cancer.
XX
OS   Homo sapiens.
XX
PN   US2003027256-A1.
XX
PD   06-FEB-2003.
XX
PF   27-DEC-2001; 2001US-00033435.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
```

```
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-456352/43.
DR   N-PSDB; ACD07776.
XX
PT   New isolated PRO polypeptide and encoding nucleic acids, useful for the
PT   diagnosis and treatment of disorders such as inflammatory bowel disease,
PT   systemic lupus erythematosis, rheumatoid arthritis, diabetes mellitus and
PT   cancer.
XX
PS   Claim 12; Fig 18; 119pp; English.
XX
CC   The invention describes an isolated nucleic acid (I) comprising at least
CC   80 % sequence identity to a nucleotide sequence that encodes a
CC   polypeptide having any of 9 125-1029 amino acid sequences (S1), or
CC   comprises at least 80 % sequence identity to or the full-length coding
CC   sequence of any of 9 662-3554 base pair sequences (S2), given in the
CC   specification. The methods and compositions of the present invention are
CC   useful for the diagnosis and treatment of disorders associated with the
CC   PRO polypeptides, such as inflammatory bowel disease, systemic lupus
CC   erythematosis, rheumatoid arthritis, systemic sclerosis, Sjogren's
CC   syndrome, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,
CC   multiple sclerosis, hepatitis, erythema multiforme, contact dermatitis,
CC   graft-versus-host-disease and cancer. This is the amino acid sequence of
CC   a novel human secreted and transmembrane PRO polypeptide
XX
SQ   Sequence 548 AA;

  Query Match             100.0%;  Score 2868;  DB 7;  Length 548;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 13
AAE38828
ID   AAE38828 standard; protein; 548 AA.
XX
AC   AAE38828;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human PRO1927 protein.
XX
KW   Secreted and transmembrane protein; tissue typing; medical application;
KW   biotechnological application; genetic disorder; industrial application;
KW   transgenic; vaccine; transgenic animal; gene therapy; human; PRO.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Signal_peptide
FT   Modified-site   5. .9
FT                   /note= "N-glycosylation site"
FT   Modified-site   6. .12
FT                   /note= "N-myristoylation site"
FT   Protein         24. .548
FT                   /note= "Mature human PRO protein"
FT   Modified-site   87. .91
FT                   /note= "N-glycosylation site"
FT   Modified-site   103. .107
FT                   /note= "N-glycosylation site"
FT   Modified-site   136. .142
FT                   /note= "N-myristoylation site"
FT   Modified-site   370. .376
FT                   /note= "N-myristoylation site"
FT   Modified-site   465. .469
FT                   /note= "N-glycosylation site"
FT   Modified-site   509. .515
FT                   /note= "N-myristoylation site"
XX
PN   US2003077657-A1.
XX
PD   24-APR-2003.
XX
PF   27-DEC-2001; 2001US-00033396.
XX
PR   04-AUG-1998;   98US-0095325P.
PR   16-DEC-1998;   98US-0112851P.
PR   16-DEC-1998;   98US-0113145P.
PR   22-DEC-1998;   98US-0113511P.
PR   12-JAN-1999;   99US-0115558P.
PR   12-JAN-1999;   99US-0115565P.
PR   12-JAN-1999;   99US-0115733P.
PR   09-FEB-1999;   99US-0119341P.
PR   10-FEB-1999;   99US-0119537P.
PR   12-FEB-1999;   99US-0119965P.
PR   02-JUN-1999;   99WO-US012252.
PR   29-OCT-1999;   99US-0162506P.
PR   01-DEC-1999;   99WO-US028634.
PR   02-DEC-1999;   99WO-US028551.
```

```
PR   09-DEC-1999;   99US-0170262P.
PR   11-FEB-2000; 2000WO-US003565.
PR   22-FEB-2000; 2000WO-US004414.
PR   02-MAR-2000; 2000WO-US005841.
PR   03-MAR-2000; 2000US-0187202P.
PR   30-MAR-2000; 2000WO-US008439.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   01-DEC-2000; 2000WO-US032678.
PR   25-MAY-2001; 2001US-00866034.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Desnoyers L,  Ferrara N,  Fong S,  Gao W,  Goddard A;
PI   Gurney AL,  Pan J,  Roy MA,  Stewart TA,  Tumas D,  Watanabe CK;
PI   Wood WI;
XX
DR   WPI; 2003-635077/60.
DR   N-PSDB; AAD58952.
XX
PT   Isolated secreted and transmembrane PRO polypeptides e.g. PRO3434 and
PT   PRO1927, useful in the preparation of a medicament for treating a
PT   condition responsive to PRO polypeptide, and as therapeutic agents e.g.
PT   vaccines.
XX
PS   Claim 12; Fig 18; 125pp; English.
XX
CC   The invention relates to secreted and transmembrane polypeptides
CC   designated as PRO (e.g. PRO1800, PRO539, PRO982, PRO1434, PRO1863,
CC   PRO1917, PRO1868, PRO3434 and PRO1927) and nucleic acid molecules
CC   encoding such polypeptides. Sequences of the invention are useful in
CC   tissue typing, gene therapy and in the preparation of vaccines.
CC   Polypeptides of the invention are useful as molecular weight markers for
CC   protein electrophoresis, as therapeutic agent for in vivo therapeutic
CC   purposes and for screening compounds that modulate their activity. They
CC   are also useful in biotechnological, industrial and medical applications.
CC   Polynucleotides of the invention are used for constructing hybridisation
CC   probes for mapping the gene encoding PRO and for the genetic analysis of
CC   individuals with genetic disorders. They are also useful for generating
CC   transgenic animals or knockout animals for the development and screening
CC   of therapeutically useful reagents. The present sequence is human PRO
CC   protein
XX
SQ   Sequence 548 AA;

  Query Match            100.0%;  Score 2868;  DB 7;  Length 548;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-270;
  Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548


RESULT 14
AAW63559
ID   AAW63559 standard; protein; 548 AA.
XX
AC   AAW63559;
XX
DT   24-NOV-1998  (first entry)
XX
DE   Human beta(1 -> 4)-N-acetylglucosaminyl-transferase (GnT-IV)b enzyme.
XX
KW   Beta(1 -> 4)-N-acetylglucosaminyl-transferase; GnT-IV; bovine; human;
KW   enzyme; sugar chain subunit; branched oligosaccharide; polysaccharide;
KW   drug; reagent; food; biopolymer; glycoprotein; erythropoietin.
XX
OS   Homo sapiens.
XX
PN   WO9826053-A1.
XX
PD   18-JUN-1998.
XX
PF   10-DEC-1997;   97WO-JP004546.
XX
PR   12-DEC-1996;   96JP-00332411.
PR   18-JUN-1997;   97JP-00161462.
XX
PA   (KIRI ) KIRIN BEER KK.
XX
PI   Oguri S,  Minowa M,  Yoshida A,  Taniguchi N,  Takeuchi M;
XX
DR   WPI; 1998-348516/30.
DR   N-PSDB; AAV38385.
XX
PT   Recombinant beta(1-4)N-acetyl:glucosaminyl:transferase - allows
PT   production of difficultly accessible branched poly:saccharides for food
PT   and drug use.
XX
PS   Claim 9; Page 70-74; 112pp; Japanese.
XX
CC   This represents a human beta(1 -> 4)-N-acetylglucosaminyl-transferase
CC   (GnT-IV)b enzyme. The invention provides bovine and human GnT-IV enzymes
CC   that can be used for converting sugar chain subunits having one structure
CC   to another structure. Vectors containing the DNA sequences encoding these
CC   enzymes can be used to transform host cells for the production of the GnT
CC   -IV enzymes. The enzymes are useful in the production of branched
CC   oligosaccharides and polysaccharides which are difficult of access by
CC   other methods. They are also useful in the production of drugs, reagents
CC   and foods and in modifying the properties of biopolymers containing sugar
CC   chains. The enzyme may also be used for the preparation of glycoproteins
CC   such as erythropoietin
XX
SQ   Sequence 548 AA;

  Query Match            99.9%;  Score 2865;  DB 2;  Length 548;
  Best Local Similarity  99.8%;  Pred. No. 2.2e-270;
  Matches  547;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDRLHAAEQESLKR 60

Qy         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHHLP 120

Qy        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HLLAKESSLQPAVRVGQGRTGVSVVMGIPSVRREVHSYLTDTLHSLISELSPQEKEDSVI 180

Qy        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VVLIAETDSQYTSAVTENIKALFPTEIHSGLLEVISPSPHFYPDFSRLRESFGDPKERVR 240

Qy        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILEFSQL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        241 WRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMILDFSQL 300

Qy        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIR 360

Qy        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 FKPSLFQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHFTLEKAYL 420

Qy        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 REDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLPFDNPQSDK 480

Qy        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 EALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGPLEALRLSIQTDSPVWVILS 540

Qy        541 EIFLKKAD 548
              ||||||||
Db        541 EIFLKKAD 548
```

```
RESULT 15
AAB94456
ID   AAB94456 standard; protein; 563 AA.
XX
AC   AAB94456;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:15104.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 15104; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 563 AA;
```

```
  Query Match             94.2%;  Score 2701;  DB 4;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-254;
  Matches  516;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         33 GDVVDVYQREFLALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGR 92
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         48 GDVVDVYQREFLALRDRLHAAEQESLKRSKELNLVLDEIKRAVSERQALRDGDGNRTWGR 107

Qy         93 LTEDPRLKPWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVR 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        108 LTEDPRLKPWNGSHRHVLHLPTVFHHLPHLLAKESSLQPAVRVGQGRTGVSVVMGIPSVR 167

Qy        153 REVHSYLTDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLL 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 REVHSYLTDTLHSLISELSPQEKEDSVIVVLIAETDSQYTSAVTENIKALFPTEIHSGLL 227

Qy        213 EVISPSPHFYPDFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVA 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 EVISPSPHFYPDFSRLRESFGDPKERVRWRTKQNLDYCFLMMYAQSKGIYYVQLEDDIVA 287

Qy        273 KPNYLSTMKNFALQQPSEDWMILEFSQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLL 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 KPNYLSTMKNFALQQPSEDWMILEFSQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLL 347

Qy        333 DHILWVKVCNPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQAL 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 DHILWVKVCNPEKDAKHCDRQKANLRIRFKPSLFQHVGTHSSLAGKIQKLKDKDFGKQAL 407

Qy        393 RKEHVNPPAEVSTSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFR 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 RKEHVNPPAEVSTSLKTYQHFTLEKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFR 467

Qy        453 SGNIEHPEDKLFNTSVEVLPFDNPQSDKEALQEGRTATLRYPRSPDGYLQIGSFYKGVAE 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 SGNIEHPEDKLFNTSVEVLPFDNPQSDKEALQEGRTATLRYPRSPDGYLQIGSFYKGVAE 527

Qy        513 GEVDPAFGPLEALRLSIQTDSPVWVILSEIFLKKAD 548
              ||||||||||||||||||||||||||||||||||||
Db        528 GEVDPAFGPLEALRLSIQTDSPVWVILSEIFLKKAD 563
```

Search completed: July 20, 2004, 10:10:47
Job time : 57 secs